REU 2013: Apprentice Program **Summer 2013**

Lecture 15: July 22, 2013

Madhur Tulsiani Scribe: David Kim

1 Expander Mixing Lemma

Let $G = (V, E)$ be d-regular, such that its adjacency matrix A has eigenvalues $\mu_1 = d \geq \mu_2 \geq \cdots \geq \mu_n$. Let $S, T \subseteq V$, not necessarily disjoint, and $E(S, T) = \{s \in S, t \in T : \{s, t\} \in E\}$.

Question: How close is $|E(S, T)|$ to what is "expected"?

Exercise 1.1 *Given G as above, if one picks a random pair $i, j \in V(G)$, what is $\mathbb{P}[i, j$ are adjacent$]$?*
Answer: $\mathbb{P}[\{i, j\} \in E] = \frac{\#good\ cases}{\#all\ cases} = \frac{\#edges}{\binom{n}{2}} = \frac{\frac{dn}{2}}{\binom{n}{2}} = \frac{d}{n-1}$

Since for $S, T \subseteq V$, the number of possible pairs of vertices from S and T is $|S| \cdot |T|$, we "expect" $|S| \cdot |T| \cdot \frac{d}{n-1}$ out of those pairs to have edges.

Exercise 1.2 ** $||E(S, T)| - |S| \cdot |T| \cdot \frac{d}{n}| \leq \mu\sqrt{|S| \cdot |T|}$*

Note that the smaller the value of μ, the closer $|E(S, T)|$ is to what is "expected", hence expander graphs are sometimes called pseudorandom graphs for μ small.

We will now show a generalization of the Spectral Theorem, with a generalized notion of eigenvalues to matrices that are not necessarily square.

2 Singular Value Decomposition

Recall that by the Spectral Theorem, given a normal matrix $A \in \mathbb{C}^{n \times n}$ $(A^*A = AA^*)$, we have

$$A = UDU^* = \sum_{i=1}^{n} \lambda_i u_i u_i^*$$

where U is unitary, D is diagonal, and we have a decomposition of A into a nice orthonormal basis, consisting of the columns of U. We show that for any rectangular matrix $A \in \mathbb{C}^{m \times n}$, we can also have a decomposition in the following form

$$A = U\Sigma V^*$$

where $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ are unitary (orthogonal if real), and Σ is "almost diagonal", giving us a nice pair of orthonormal bases to work with.

Definition 2.1 (Singular Value) *$\sigma \geq 0$ is called a "singular value" if $\exists u \in \mathbb{C}^m, v \in \mathbb{C}^n$ $(||u|| = ||v|| = 1)$ such that $Av = \sigma u$ and $A^*u = \sigma v$, i.e. u and v are like "pairs". v is called "the right singular vector", and u is called "the left singular vector".*

Exercise 2.2 *(σ is a singular value of A) $\Leftrightarrow$ (σ^2 is an eigenvalue of AA^* and A^*A).*

($\Rightarrow$) By assumption, $(\exists u, v)(Av = \sigma u, A^*u = \sigma v)$. Then $AA^*u = A(\sigma v) = \sigma Av = \sigma^2 u$. Likewise, $A^*Av = A^*(\sigma u) = \sigma A^*u = \sigma^2 v$.
($\Leftarrow$) Before proving the converse, what can we infer about AA^*, A^*A and σ^2 assuming that σ^2 is an eigenvalue?

Recall that a matrix $M \in \mathbb{C}^{n\times n}$ is *Hermitian* if $M = M^*$. We proved in a previous lecture that the eigenvalues of a Hermitian matrix are real.

Clearly AA^* and A^*A are Hermitian, so $\sigma^2 \in \mathbb{R}$. Also, $(\forall x \in \mathbb{C}^m)(x^*AA^*x = \langle A^*x, A^*x\rangle \geq 0) \Rightarrow AA^*$ is positive semidefinite $\Rightarrow \sigma^2 \geq 0$, as σ^2 is an eigenvalue. The next exercise shows that given such σ^2, we can assume that $\sigma \geq 0$.

Exercise 2.3 *WLOG, all singular values $\sigma \geq 0$.*

Proof: If $\sigma < 0$, let $\sigma' = -\sigma > 0$. Then $Av = \sigma u = (-\sigma)(-u) = \sigma'(-u)$, and $A^*u = \sigma v \Rightarrow A^*(-u) = -\sigma v = \sigma' v$. ∎

For the rest of this lecture, we will stick to real matrices, but we can generalize the same results to complex matrices. Recall we defined $B \in \mathbb{R}^{n\times n}$, $B = B^T$, as PSD (positive semi-definite) if $(\forall x \in \mathbb{R}^n)(x^TBx \geq 0)$. The same definition follows for complex matrices: $B \in \mathbb{C}^{n\times n}$, $B = B^*$, is PSD if $(\forall x \in \mathbb{C}^n)(\langle x, Bx\rangle \geq 0)$.

We now prove the main theorem:

Theorem 2.4 (Singular Value Decomposition) *Let $A \in \mathbb{R}^{m\times n}$ such that $rk(A) = r$. Then $\exists \sigma_1 \geq \cdots \geq \sigma_r > 0$ and an orthogonal matrix $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}, \Sigma \in \mathbb{R}^{m\times n}$ such that $A = U\Sigma V^T$, where $\Sigma_{ii} = \sigma_i$ for $i = 1, \ldots, r$, and $\Sigma_{ij} = 0$ everywhere else.*

Proof strategy: We will use the Spectral Theorem to decompose $A^TA = VDV^T$, and take the diagonals of D as σ_i^2, the eigenvalues.
Question: Why is $\sigma_r > 0$?
Answer: By the following exercise, $rk(A) = rk(A^TA) = r$, and $A^TA \sim D$ with σ_i^2's on the diagonal.

Exercise 2.5 *Prove: $(\forall A \in \mathbb{R}^{m\times n})(rk(A^TA) = rk(A))$.*
hint: rank-nullity theorem.

Proof: [Singular Value Decomposition]

A^TA is symmetric - hence by the Spectral Theorem, $A^TA = VDV^T$ where $D \in \mathbb{R}^{n\times n}$ is diagonal with non-negative entries, and V is orthogonal. The rank of D is r, so it has r positive entries on the diagonal, say $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r > 0$ on the first r diagonal entries, and 0's everywhere else.

We simply use the same V above (the eigenbasis of A^TA), and see what we get from AV. The claim is that this gives us the desired decomposition.

- Let $V = \left[\begin{array}{c|c} V_1 & V_2 \end{array}\right] \in \mathbb{R}^{n\times n}$, where V_1 consists of the first r columns , and V_2 the last $n-r$ columns of V.
- Let $U = \left[\begin{array}{c|c} U_1 & U_2 \end{array}\right] \in \mathbb{R}^{m\times m}$, where $U_1 = AV_1(D')^{-1/2} \in \mathbb{R}^{m\times r}$ and $(D')^{-1/2} \in \mathbb{R}^{r\times r}$ with ${(D')^{-1/2}}_{ii} = \frac{1}{\sqrt{\lambda_i}}$. Let U_2 be any extension of the remaining $m-r$ columns from U_1 to an orthonormal basis.
- Let $\Sigma = \left[\begin{array}{c|c} (D')^{1/2} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \in \mathbb{R}^{m\times n}$, $(D')^{1/2} \in \mathbb{R}^{r\times r}$ where ${(D')^{1/2}}_{ii} = \sqrt{\lambda_i}$

Then $U_1 = \left[\frac{Av_1}{\sqrt{\lambda_1}} \dots \frac{Av_r}{\sqrt{\lambda_r}}\right]$. U would be orthogonal if U_1 consists of othonormal columns.

Claim 2.6 *The columns of U_1 are othonormal.*
Proof. Since V is an orthogonal matrix,

$$\begin{aligned} \langle u_i, u_j \rangle = \left\langle \frac{Av_i}{\sqrt{\lambda_i}}, \frac{Av_j}{\sqrt{\lambda_j}} \right\rangle &= \frac{v_i^T(A^TAv_j)}{\sqrt{\lambda_i\lambda_j}} = \frac{\lambda_j v_i^T v_j}{\sqrt{\lambda_i\lambda_j}} \\ &= \frac{\lambda_j \langle v_i, v_j \rangle}{\sqrt{\lambda_i\lambda_j}} = \begin{cases} 0 & \textit{if } i \neq j \\ 1 & \textit{if } i = j \end{cases} \end{aligned}$$

Thus, we can extend the columns of U_1 with U_2 such that U is orthogonal. To complete the proof, we only need to prove that such U, V, Σ satisfy the properties we desire.

Claim 2.7 *Given A, V, U, Σ as above, $A = U\Sigma V^T$.*
Proof.

$$\begin{aligned} U\Sigma V^T &= \left[\begin{array}{c|c} U_1 & U_2 \end{array}\right] \left[\begin{array}{c|c} (D')^{1/2} & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} V_1^T \\ \hline V_2^T \end{array}\right] \\ &= \left[\begin{array}{c|c} U_1 & U_2 \end{array}\right] \left[\begin{array}{c} (D')^{1/2}V_1^T \\ \hline 0 \end{array}\right] \\ &= U_1(D')^{1/2}V_1^T = AV_1(D')^{-1/2}(D')^{1/2}V_1^T \\ &= AV_1V_1^T = AI \\ &= A \end{aligned}$$

■

Now that we have the decomposition, we check that the decomposition indeed gives us the singular values, which would complete the proof for Exercise 2.2.

Exercise 2.8 *If $A = U\Sigma V^T$, where $rk(A) = rk(\Sigma) = r$, then $\forall i = 1 \dots r$, where $\sigma_i = \Sigma_{ii}$,*

$$Av_i = \sigma_i u_i$$
$$A^T u_i = \sigma_i v_i$$

Proof:

$$Av_i = U\Sigma V^T v_i = U\Sigma \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \sigma_i u_i$$

(Likewise for $A^T u_i = \sigma_i v_i$.) ■

Exercise 2.9 *If $A = U\Sigma V^T \Rightarrow A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$, where u_i and v_i are i'th columns of U and V, respectively.*

3 Extension of Courant-Fischer

Now we have the SVD theorem. We will show that singular values also have an analogous definition similar to that of the Courant-Fischer for eigenvalues.

Exercise 3.1 *Show that for $A \in \mathbb{R}^{m\times n}$, the k'th largest singular value of A is*

$$\sigma_k = \underset{\substack{S \subseteq \mathbb{R}^n \\ \dim(S) = n-k+1}}{min} \; \underset{x \in S, x \neq 0}{max} \frac{||Ax||}{||x||} = \underset{\substack{T \subseteq \mathbb{R}^n \\ \dim(T) = k}}{max} \; \underset{x \in T, x \neq 0}{min} \frac{||Ax||}{||x||}$$

hint: Use Courant-Fischer for $A^T A$.

Proof: Let's show for $k = 1$. (rest is exercise)

$$\sigma_1 = \underset{x \in \mathbb{R}^n, x \neq 0}{max} \frac{||Ax||}{||x||} \Leftrightarrow \sigma_1^2 = \underset{x \in \mathbb{R}^n, x \neq 0}{max} \frac{||Ax||^2}{||x||^2}$$
$$= \underset{x \in \mathbb{R}^n, x \neq 0}{max} \frac{x^T A^T A x}{x^T x}$$
$$= \underset{x \in \mathbb{R}^n, x \neq 0}{max} R_{A^T A}$$

We know that the right side is true from Courant-Fischer, as σ_1^2 is the largest eigenvalue of $A^T A$. Question: what is the x that achieves σ_1? v_1, the first column of V in $A = U\Sigma V^T$. ■

4 Applications

4.1 Data Fitting

The interpretation of singular values in extension of Courant-Fischer has many fairly good applications, and a nice interpretation in terms of fitting data to a line, or a subspace.

Suppose we have m points in $\mathbb{R}^n$, and we wish to find a line passing through the origin "fitting the data" the best. By "fitting the data" the best, we mean to find the line l which minimizes $\sum_{i=1}^{m}(d_i(l))^2$ (least squared distance), where $d_i(l)$ = closest point of l to the i'th data point.

How do you specify any line passing through the origin? a unit vector in the direction of the line. So let v be the unit vector in the direction of a line l.

Question: Let $a_i \in \mathbb{R}^n$, and l in direction of v, a unit vector. What is $d_i(l)$?
Answer: By a simple geometric argument,

$$||a_i||^2 = |\langle a_i, v\rangle|^2 + (d_i(l))^2 \Rightarrow (d_i(l))^2 = ||a_i||^2 - |\langle a_i, v\rangle|^2$$

Since $||a_i||^2$ does not depend on v,

$$\begin{aligned} min\sum_{i=1}^{m}(d_i(l))^2 &= \min_{||v||=1, v\in\mathbb{R}^n}\sum_{i=1}^{m}(||a_i||^2 - |\langle a_i, v\rangle|^2) \\ &= \sum_{i=1}^{m}||a_i||^2 - \max_{||v||=1, v\in\mathbb{R}^n}\sum_{i=1}^{m}|\langle a_i, v\rangle|^2 \\ &= \sum_{i=1}^{m}||a_i||^2 - \max_{||v||=1, v\in\mathbb{R}^n}||Av||^2 \end{aligned}$$

where $A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix}$. We already solved this maximization problem: it is the first right singular vector, v, of A in the singular value decomposition.

We can generalize the problem to finding $S, \dim(S) = k$ which maximizes $\sum_{i=1}^{m}(d_i(S))^2$ for m points in $\mathbb{R}^n$, where $d_i(S)$ is the closest point from a point to the subspace. (In general, given a_i, take its projection to S, obtained by innerproducts with an orthnormal basis of S. Then subtract the projection from a_i - the length of the resulting vector is $d_i(l)$.)

Exercise 4.1 *In the above setting, prove that $S = span(v_1, \ldots, v_k)$ is the subspace minimizing $\sum_i(d_i(l))^2$, where v_i are the right singular vectors of A.*

4.2 Approximating a Matrix

This leads to our final application, where SVD is most useful - approximating a matrix.

Problem: Given $A \in \mathbb{R}^{m \times n}$, find the "best" rank k matrix B which minimizes $\sum_{i,j}(A_{ij} - B_{ij})^2 = ||A - B||_F^2$.

Definition 4.2 (Frobenius Norm) $||M||_F^2 = \sum_{i,j}(M_{ij})^2$

Claim 4.3 *Let $A = U\Sigma V^T = \sum_{i=1}^r \sigma_i u_i v_i^T$. The best rank k approximation of A is $A_k = U\Sigma' V^T$, where Σ' is the matrix which keeps only the first k singular values of Σ, and is 0 everywhere else, i.e. $A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$.*

Proof: We will prove $||A - A_k||_F^2 = \min\limits_{rank(B)=k} ||A - B||_F^2$. Let's first see what B should look like:

$$||A - B||_F^2 = \sum_{i=1}^m ||A_i - B_i||^2$$

where A_i, B_i are the i'th rows of A and B, respectively. Let $W := row - space(B)$. Then $\dim(W) = k$.

Question: Given W, what is the best B_i which minimizes $||A_i - B_i||^2$? The projection of A_i to W. Then we know that for each i,

$$||A_i - B_i||^2 \geq ||A_i - proj(A_i, W)||^2$$

and we have

$$\sum_i ||A_i - B_i||^2 \geq \sum_i ||A_i - proj(A_i, W)||^2$$

We already know that the first k singular vectors minimize the this. ■

Exercise 4.4 *Let $A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$. Then $\forall i, (A_k)_i = proj(A_i, span(v_1, \ldots, v_k))$.*

How do we calculate the projection of A_i to $span(v_1, \ldots, v_k)$? Since V is orthogonal, simply take $\sum_{j=1}^k \langle A_i, v_j \rangle v_j$. The above exercise asks you to show that this is exactly equal to the i'th row of A_k.