RAYLEIGH QUOTIENT AND THE MIN-MAX THEOREM

Definition 1. A Matrix is called *Hermitian* if $A^\star = A$. Notice that if $A = A^\star$ then $A^\star A = A^\star A$ so by the diagonalization theorem for normal matrices, A is unitarily diagonalize i.e. it has an orthonormal basis of eigenvectors.

Lemma 2. *Let A be Hermitian and suppose its Eigenvalues are $\lambda_1 \leq \ldots \leq \lambda_n$ with corresponding eigenvectors $v_1, \ldots, v_n$. If S_k is a $k-$dim'l subspace then:*

$$\exists x \in S_k \; s.t. \; \frac{\langle Ax, x\rangle}{\langle x, x\rangle} \geq \lambda_k$$

Proof. By a dimension counting argument, $S_k \cap \text{span}\{v_k, \ldots, v_n\} \neq \phi$ so there exists $x \in S_k$ with $x \in \text{span}\{v_k, \ldots, v_n\}$. For this x it is easy to verify $\frac{\langle Ax, x\rangle}{\langle x, x\rangle} \geq \lambda_k$. □

Theorem 3. *(The Min-Max Theorem) Let A be Hermitian and suppose its Eigenvalues are $\lambda_1 \leq \ldots \leq \lambda_n$:*

$$\min_{dim S_k = k} \max_{x \in S_k} \frac{\langle Ax, x\rangle}{\langle x, x\rangle} = \lambda_k$$

Proof. By the above lemma, the LHS is $\geq \lambda_k$. Choosing $S_k = \text{span}\{v_k, \ldots, v_n\}$ gives LHS $\leq \lambda_k$. □

Definition 4. A Hermitian matrix is called positive definite if all of its eigenvales are > 0. Equivalently, if $\langle Ax, x\rangle > 0$ for all non-zero x.

Theorem 5. *Say M, L are Hermitian with eigenvalues $\mu_1 \leq, \ldots, \leq \mu_n$ and $\lambda_1 \leq, \ldots \leq \lambda_n$ respectively. If $L - M$ is positive definite (i.e "$L > M$") then $\lambda_k > \mu_k$ for each k.*

Proof. (Using Rayleigh Quotient) Write:

$$\min_{dim S_k = k} \; \max_{x \in S_k} \tfrac{\langle Lx, x\rangle}{\langle x, x\rangle} = \lambda_k$$
$$\min_{dim S_k = k} \; \max_{x \in S_k} \tfrac{\langle Mx, x\rangle}{\langle x, x\rangle} = \mu_k$$

Let T be the k-dimensional subspace that achieves the minimum for the λ_k above and let $x_0 \in T$ be the vector that achieves $\max_{x \in S_k} \frac{\langle Mx, x\rangle}{\langle x, x\rangle}$. Have then (using $\langle Lx, x\rangle > \langle Mx, x\rangle$ for every x):

$$\begin{aligned} \lambda_k &= \min_{dim S_k = k} \max_{x \in S_k} \frac{\langle Lx, x\rangle}{\langle x, x\rangle} \\ &= \max_{x \in T} \frac{\langle Lx, x\rangle}{\langle x, x\rangle} \\ &> \max_{x \in T} \frac{\langle Mx, x\rangle}{\langle x, x\rangle} \\ &= \frac{\langle Mx_0, x_0\rangle}{\langle x_0, x_0\rangle} \end{aligned}$$

On the other hand:

$$\begin{aligned} \mu_k &= \min_{dim S_k = k} \max_{x \in S_k} \frac{\langle Mx, x\rangle}{\langle x, x\rangle} \\ &\leq \max_{x \in T} \frac{\langle Mx, x\rangle}{\langle x, x\rangle} \\ &= \frac{\langle Mx_0, x_0\rangle}{\langle x_0, x_0\rangle} \end{aligned}$$

So finally we have $\lambda_k > \frac{\langle Mx_0, x_0\rangle}{\langle x_0, x_0\rangle} \geq \mu_k$ □

Definition 6. We say M is positive definite if M is hermitian and all eigenvalues are ≥ 0. Equiivalently $\langle Mx, x\rangle > 0$ for all x.

1. SVD decomopisition

Hermitian Matices are very nice to work with because they have:

- An orthonormal set of eigenvectors
- Real valued eigenvalues

The SVD decomoposition is an attempt to get as much of these properties as we can onto an ordinary (non-Hermitian) matrix A. This works even from rectangular matrices! We first need the following lemma:

Lemma 7. *If A is an $n \times m$ matrix and B is an $m \times n$ matrix, then AB and BA have the same eigenvalues except for 0. More precisly: Eigenvalues$(AB) \cup \{0\} =$ Eigenvalues$(BA) \cup \{0\}$.*

Proof. One proof is by using Sylvesters determinant identity $\det(I_n + AB) = \det(I_m + BA)$. Another proof is more hands on. We show inclusions both ways. Suppose v is an eigenvector of AB of eigenvalue λ. If $\lambda = 0$ then $\lambda \in$ Eigenvalues$(BA) \cup \{0\}$ and we are done. Otherwise $\lambda \neq 0$ and we have $ABv = \lambda v \implies (BA)Bv = \lambda Bv$. Notice moreover that $Bv \neq 0$ (if this were the case, λ would have to be zero). Hence Bv is an eigenvector of BA of eigenvalue λ. Hence $\lambda \in$ Eigenvalues$(BA) \cup \{0\}$. □

Lemma 8. *(Assume $n > m$) Moreover, for diagonolizable matrices, the multiplicities of the eigenvalues are the same too (execpt for zero): $\dim null(AB - \lambda) = \dim null(BA - \lambda)$ for $\lambda \neq 0$ and for $\lambda = 0$, $\dim null(AB) = \dim null(BA) + (n - m)$ to make up the difference.*

Proof. For $\lambda \neq 0$, let $v_1, \ldots, v_k$ be a basis for null$(AB - \lambda)$, i.e. eigenvectors of eigenvalue λ. By the above, each Bv_i is an eigenvector of BA of eigenvalue λ. Moreover, the vectors $Bv_1, \ldots, Bv_k$ are indepedent, because otherwise $\sum \alpha_j B(v_j) = B(\sum \alpha_j v_j) = 0 \implies \sum \alpha_j v_j \in$ null(AB). But $\sum \alpha_j v_j$ is an eigenvector of eigenvalue λ so this can only happen $\sum \alpha_j v_j = 0$ so all the α_j must be zero! This shows dimnull$(AB - \lambda) \leq$ dimnull$(BA - \lambda)$, and the reverse inequality □

Theorem 9. *Let A be an $m \times n$ matrix. Assume WOLOG $n > m$. Notice that AA^T is a Hermitian $m \times m$ matrix with non-negative eigenvalues (why?). Therefore, there is an orthonormal set of eigenvectors $\{u_1, \ldots, u_m\}$ and non-negative eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. Similarly, $A^T A$ is a Hermitian $n \times n$ matrix with non-negative eigenvalues. Call its eigenvectors $\{v_1, \ldots, v_n\}$ and its eigenvalues $\lambda_1, \ldots, \lambda_m, 0, 0, \ldots 0$. (Since AA^T and $A^T A$ have the same eigenvalues except for 0, we already have the eigenvalues of $A^T A$ labeled!)*

*The square roots $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$ are called the **singular values of** A and the **singular value decomposition is:***

$$\underbrace{A}_{m\times n} = \underbrace{\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_m \\ | & | & & | \end{bmatrix}}_{m\times m} \underbrace{\begin{bmatrix} \sqrt{\lambda_1} & 0 & & 0 & | & | & & | \\ 0 & \sqrt{\lambda_2} & & 0 & | & | & & | \\ & & \ddots & & 0 & 0 & \dots & 0 \\ & & & \dots & | & | & & | \\ 0 & 0 & & \sqrt{\lambda_m} & | & | & & | \end{bmatrix}}_{m\times n} \underbrace{\begin{bmatrix} - & v_1 & - \\ - & v_2 & - \\ & \vdots & \\ - & v_n & - \end{bmatrix}}_{n\times n}$$

Proof. (For simplicity we assume that $\lambda_1, \ldots, \lambda_m$ are distinct. The general proof is not much different). The above construction gives us three ingredients: An o.n.b $\{u_1, \ldots, u_m\}$ for $\mathbb{R}^m$, An o.n.b $\{v_1, \ldots, v_n\}$ for $\mathbb{R}^n$, and the eigenvalues $\lambda_1, \ldots, \lambda_m$. Since A is a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$, A is completly characterized by its action on the basis $\{v_1, \ldots, v_n\}$. The output can be probed by inner products $\langle \cdot, u_i \rangle$ since these are an o.n.b. In other words, its suffices to check that $\langle Av_i, u_j \rangle$ is the same as the action of the above decomposition. If v $\lambda_i \neq 0$, then notice that Av_i is a an eigenvector of AA^T with eigenvalue λ_i by the following argument:

$$(AA^T)(Av_i) = A(A^T A)v_i = A(\lambda_i v_i) = \lambda_i (Av_i)$$

But in this case, Av_i must be in the eigenspace of λ_i for the matrix AA^T! In other words, Av_i is perpendicular to all u's except u_i. Moreover, $Av_i \in$ span(u_i) and we can find the constant $Av_i = cu_i$ as follows: $c = \langle Av_i, u_i \rangle = \langle v_i, A^T u_i \rangle = \langle v_i, \frac{\lambda_i}{c} v_i \rangle = \frac{\lambda_i}{c} \implies c = \sqrt{\lambda_i}$ (we used $Av_i = cu_i \implies \lambda_i v_i = A^T A v_i = cA^T u$ here)

On the other hand, its easy to check that the above decomopisiton also has $\langle UDVv_i, u_j \rangle = \sqrt{\lambda_i}\delta_{i=j}$. □

Remark 10. Using "dirac" notation for vectors, this says the matrix A can be written as:

$$A = \sum \sqrt{\lambda_i} \, |u_i\rangle \langle v_i|$$

i.e. A eats in the v_i component, stretches it by $\sqrt{\lambda_i}$ and outputs this in the direction u_i.