1 Eigenvalues and Eigenvectors

Some of the results in this chapter deal with general linear operators. Most, however, will be about matrices with complex entries. Many of the concepts covered will be similar to a college linear algebra course (eigenvalues, eigenvectors, diagonalization, ect.), but with a more of the proofs presented. These notes mostly go along with the book, and in some places refer to the book. Some of the proofs are a little different or use different wording.

Definition 1.1. A function $F : A \to A$ is a linear operator iff F preserves addition and scaler multiplication.

Definition 1.2. Suppose $L : V \to V$ is a linear operator and v is a finite dim vector space over $\mathbb{C}$. Then scaler $\lambda \in \mathbb{C}$ is an eigenvalue of L if there is an $x \in (V \setminus \{\mathbf{0}\})$ such that,

$$L(x) = \lambda \cdot x \tag{1}$$

Definition 1.3. Any $x \in (V \setminus \{\mathbf{0}\})$ that satisfies (1) for some eigenvalue x is an eigenvector.

Definition 1.4. For any $\lambda \in \mathbb{C}$ define the λ-eigenspace of L to be,

$$\Sigma_\lambda(L) = \{x \in V | L(x) = \lambda \cdot x\} \tag{2}$$

(Note that this will contain the zero vector for all λ even though it is not an eigenvector)

Remark. Note that,

$$\Sigma_\lambda(L) = \mathcal{N}(L - \lambda I) \tag{3}$$

(I.e. the eigenspace of λ is the null space of $(L - \lambda I)$) Therefor anything in the null-space of $(L - \lambda I)$ is either an eigenvector to L or the zero vector.

Theorem 1.1. *For $n \times n$ matrix A and scaler λ, the following are equivalent:*

1. *λ is an eigenvalue of A*
2. *$\exists\, x \neq \mathbf{0}$ s.t. $(\lambda I - A)x = \mathbf{0}$*
3. *$\Sigma_\lambda(A) \neq \{\mathbf{0}\}$*

4. $\lambda I - A$ *is singular*

5. $|\lambda I - A| = 0$

Proof. It should be simple to see that $1 \Leftrightarrow 2$, $2 \Leftrightarrow 3$ and $2 \Leftrightarrow 4$, then $4 \Leftrightarrow 5$ is a well known result in linear algebra.

Definition 1.5. For $n \times n$ matrix A,

$$p_A(z) := |zI - A| \tag{4}$$

is the characteristic polynomial of A.

Proposition 1.2. *For $n \times n$ matrix A, the characteristic polinomial can be factored over the complex numbers to be,*

$$p(z) = \prod_{j=1}^{r} (z - \lambda_j)^{m_j} \tag{5}$$

Where λ_j are the distinct roots of $p(z)$ and $\sum_j^r m_j = n$.

Proof. From the fundamental theorem of algebra we know that any "monic" polynomial for which the term of the highest degree has a multiple of one (e.g. $x^n + yx^{n-1} + zx^{n-2} + ...$) can be decomposed into a product as in (5). So we just need to show that the determinate of $(zI - A)$ is always a n-degree, monic polynomial. For this we think about the cofactor expansion of the determinate along any row with all the zs in the diagonals (always using the first row in the expansion is simplest). Then it is easy to see that the factor with the highest term in the polynomial is z^n.

Definition 1.6. Let A and B be $n \times n$ matrices. Define them to be similar matrices iff $A = P^{-1}BP$ for some nonsingular $n \times n$ matrix P.

Proposition 1.3. *If A and B are similar matrices then,*

1. $p_A(z) = p_B(z)$

2. *A and B have the same eigenvalues.*

Proof. Consider

$$\begin{aligned} p_A(z) =& |zI - A| \\ =& |zI - P^{-1}BP| \\ =& |P^{-1}(zI - B)P| \\ =& |P^{-1}| \cdot |(zI - B)| \cdot |P| \\ =& |(zI - B)| = p_B(z) \end{aligned}$$

This gives us 1. Then it must be that $1 \Rightarrow 2$ by part 5 of Theorem 1.1.

2 Invariant Subspaces

Definition 2.1. Subspace $W \subseteq V$ is invariant under L iff $L(W) \subseteq W$

Note that if we consider a function whose domain and range are the same, i.e. $L : A \to A$, A will always be invariant.

Theorem 2.1. *Let W with basis $S = \{s_1, s_2, .., s_k\}$ be L-invariant, where L is a linear operator on $V \supseteq W$. Expand the basis S to be a basis S' for the entire space V, then the square matrix representation of L in the basis S' is a block upper-triangular matrix with three non-zero block, where the top left block is the representation of L on W in basis S.* [1]

[1] From the proof on page 150, it seems to assume that L being a linear operator follows from invariance. This is certainty not the case. Consider the function $F((a_1, a_2, ...)) = (\sqrt{a_1}, \sqrt{a_2}, ...)$, this is invariant on the subspace spanned by $(1, 0, 0, ...)$, but it is not a linear operator. If you want to learn more about invariant subspaces I recommend the wikipedia page on invariant subspaces and these notes: http://math.mit.edu/ trasched/18.700.f11/lect13-beamer.pdf.

3 Diagonalization

Theorem 3.1. *Let L have k distinct eigenvalues, $\{\lambda_1, \lambda_2, ..., \lambda_k\}$, with corresponding eigenvectors, $\{x_1, x_2, ..., x_k\}$. This set of eigenvectors is linearly independent.*

Proof. Suppose not. Then it must be that $\dim(span(\{x_1, x_2, ..., x_k\})) = r < k$, and that a subset of the vectors $\{a_1, a_2, ..., a_r\} \subset \{x_1, x_2, ..., x_k\}$ will be linearly independent. Now consider the representation of one of the vectors outside of the subset in terms of the linearly independent vectors:

$$x = c_1 \cdot a_1 + c_2 \cdot a_2 + ... + c_r \cdot a_r$$

Since L is a linear operator,

$$L(x) = c_1 \cdot L(a_1) + c_2 \cdot L(a_2) + ... + c_r \cdot L(a_r)$$

or

$$\lambda_x x = c_1 \lambda_{a1} a_1 + c_2 \lambda_{a2} a_2 + ... + c_r \lambda_{ar} a_r$$

for respective eigenvalues. The first equation multiplied by λ_x subtracted from the third gives,

$$0 = c_1(\lambda_{a1} - \lambda_x)a_1 + c_2(\lambda_{a2} - \lambda_x)a_2 + ... + c_r(\lambda_{ar} - \lambda_x)a_r$$

However, since all the a vectors a linearly Independence, no linear combination of them can sum to zero, unless all constants are zero. But this would imply that $\lambda_{ia} = \lambda_x$ for all $i \in \{1, 2, ..., k\}$ so that not all the eigenvalues are distinct. This is a contradiction.

Definition 3.1. Operator L is simple if all of it's eigenvalues are distinct.

Definition 3.2. Operator L is semisimple if a basis for $\mathbb{F}^n$ can be created from a set of it's eigenvectors.

Corollary. *The eigenvectors of any $n \times n$ simple matrix span $\mathbb{F}^n$.*

Proof. An $n \times n$ simple matrix will have n distinct eigenvalues that correspond to n linearly independent eigenvectors by Theorem 3.1.

Definition 3.3. An $n \times n$ matrix is diagonalizable iff a non-singular matrix P and diagonal matrix D exist such that,

$$D = P^{-1}AP \tag{6}$$

Theorem 3.2. *An $n \times n$ matrix is diagonalizable iff it's semisimple*

Proof. ($\Rightarrow$) Let A be diagonalizable. Then (6) must hold for some diagonal matrix D and invertible matrix P. Denote columns of P, $x_1, x_2, ..., x_n$, respectively, note that these must span $\mathbb{R}^n$, since P is invertible. So now, we just need to show that these vectors are eigenvectors of A, which is shown by the equation,

$$\begin{aligned} D &= P^{-1}AP \\ PD &= AP \\ [\lambda_1 x_1, \lambda_2 x_2, ..., \lambda_n x_n] &= [Ax_1, Ax_2, ..., Ax_n] \end{aligned}$$

$\Leftarrow$ Assume A has n eigenvectors that are independent, and construct a matrix P by using the eigenvectors to create columns so, $P = [x_1, x_2, ..., x_n]$. Since P must be invertible, the following equations show that A is diagonalizable:

$$\begin{aligned} [\lambda_1 x_1, \lambda_2 x_2, ..., \lambda_n x_n] &= [Ax_1, Ax_2, ..., Ax_n] \\ PD &= AP \\ D &= P^{-1}AP \end{aligned}$$

Where D is just the diagonal matrix of eigenvectors.

Note. This process of digitalization with the eigenvectors is especially important for calculating the value of an invertible, square matrix raised to the n^{th} power. Since $A^n = (P^{-1}AP)^n = P^{-1}DPP^{-1}DP...P^{-1}DP = P^{-1}D^nP$.

Definition 3.4. Define vector x' to be a left eigenvector of A iff for some scaler λ,

$$x'A = \lambda x'$$

Note that the transpose (x) of a left eigenvector of A will be a normal "right" eigenvector of A'.

4 Shur's Lemma

Definition 4.1. A complex matrix A is orthonormal iff it satisfies,[2]

$$AA^H = A^H A = I$$

Definition 4.2. Matrices A and B are orthonormally similar iff some orthonormal matrix U exists such that

$$UAU^H = B$$

Definition 4.3. Complex matrix A is hermitian iff $A = A^H$

Lemma 4.1. *If A is hermitian and orthonormally similar to B, B is also hermitian.*

Theorem 4.2. *Every complex $n \times n$ matrix A is orthonormally similar to some upper-triangular matrix.*

Proof. The case for $n = 1$ is trivial. Now assume this hold for $n = k$ and considers $(k+1) \times (k+1)$ matrix A. Now consider the eigenvectors $x_1, x_2, ..., x_{k+1}$. We can rescale x_1 to length one, then create a orthonormal set from the rest of the eigenvectors by using the Grahm-Schmidt algorithm labeled $w_1, w_2, ..., w_{k+1}$ (only w_1, will necessarily remain an eigenvector). Define $U = [w_1, w_2, ..., w_{k+1}]$, giving us,

$$U^H AU = \text{See book pg. 158}$$

Theorem 4.3. *Let λ be an eigenvalue of operator L on $\mathbb{R}^n$. The dimensionality of the corresponding eigenspace is less than the multiplicity of the eigenvalue ($\dim \Sigma_\lambda(L) \leq m_\lambda$).*

Proof. Consider a basis of $\Sigma_\lambda(L)$, $v_1, v_2, ..., v_k$. We can choose other vectors $v_{k+1}, v_{k+2}, ..., v_n$, so that the vectors $v_1, v_2, ..., v_n$ will span all of $\mathbb{R}^n$. The matrix representation of L in the created basis will be of the form (See book pg. 159, this part follows from the definition of eigenvalues and eigenvectors)

Theorem 4.4. *Every hermitian matrix A is orthonormally similar to a diagonal matrix with all real entries.*

[2] Remember that A^H is the conjugate transpose of A. So the transpose where the sign is changed on the imaginary part.

Proof. By 4.2 we have that A is orthonormally similar to some upper triangular matrix and from Lemma 4.1, that upper triangular matrix must also be hermitian. This implies that the conjugate transpose of the upper triangular matrix is equal to itself. This can only be the case for an upper triangular matrix if it is a diagonal matrix with real entries.

Corollary. *All hermetian matrices have real eigenvalues and orthogonal eigenvectors for distinct eigenvalues.*

Proof. Think about the diagonalization result. The diagonal matrix gives the real eigenvalues and the eigenvectors are given by the columns of the orthonormal matrix.

Definition 4.4. Matrix A is normal iff $A^H A = AA^H$.

Theorem 4.5. *Matrix A is normal iff it is orthogonally diagonolizable. (I.e. iff it is orthogonally similar to a diagonal matrix with all real entries.)*

Proof. ($\Rightarrow$) Assume A is normal. By 4.2 we have that for some orthonormal U and upper triangular T, $U^H AU = T$. The following equality shows that T is normal:

$$T^H T = U^H A^H U U^H AU = U^H A^H AU = U^H AA^H U = U^H AUU^H A^H U = TT^H$$

But when the diagonal values of the LHS and RHS matrices are compared, it implies that T is diagonal (see pg. 161 for illustration).

($\Leftarrow$) This direction will follow easily from the algebra.

5 Singular Value Decomposition

Definition 5.1. An $n \times n$ matrix A is positive definite iff it is hermitian and $x^H Ax > 0$ for any $x \neq 0$ and positive semidefinite iff it is hermitian and the weak inequality holds.

Theorem 5.1. *Hermitian matrix A is positive definite (semidefinite) iff it only has positive (non-negative) eigenvalues.*

Proof. ($\Rightarrow$) Assume A is pos. def. and consider eigenvalue λ and corresponding eigenvector x. By definition, $x^H Ax = \lambda x^H x > 0$. So it must be that $\lambda > 0$.

($\Leftarrow$) Now assume that all of A's eigenvalues are positive. Since A is hermetian, we have corresponding orthonormal eigenvectors.

Theorem 5.2. *Let A be a $n \times n$ positive semidefinite with non-zero eigenvalues, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r > 0$. An orthonormal matrix Q exists such that $Q^H AQ = \text{diag}(\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r, 0, ..., 0)$, and where the first $n - r$ columns are a basis for the null space and the first n columns are a basis for a space orthogonal to the null space.*

Proof. See book. Like a lot of these proofs, the visualization of the proof will be similar to Theorem 3.2.

Proposition 5.3. *Given a positive definite matrix A, a inner product can be defined as follows $< x, y >= x^H Ay$*

Proof. An inner product has to satisfy four conditions, the first two are the linearity conditions about scalar multiplication and addition that will clearly hold from what we have shown earlier about matrix multiplication. The third condition says that $< x, x >$ is always positive, which follows from the definitions of pos def. The fourth condition symmetry, which will always hold for scalers.

Proposition 5.4. *If A is $m \times n$ with rank r, then $A^H A$ is pos semidef and has rank r.*

Proof. Matrix $A^H A$ is hermitian since $(A^H A)^H = A^H A$, and pos semidef since $x^H A^H Ax =< Ax, Ax >\geq 0$. Rank r was shown in chapter 3 of the book.

Theorem 5.5. *Singular value decomposition: let A be $m \times n$ with rank r. There exists orthonormal matrices $m \times m$ U and $n \times n$ V and $m \times n$ real, diagonal $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_r, 0, ..., 0)$ such that*

$$A = U\Sigma V^H \tag{7}$$

Proof. We know from 5.4 that $A^H A$ is positive semi-definite with rank r. From 5.2 we can decompose it into an orthonormal diagonalization (see book pg 164). Define σ_i as the positive square root of the corresponding diagonal entry.

Remark. The entries of the diagonal matrix are unique, but this is not necessarily the case for U and V.

6 Consequences of SVD

Definition 6.1. Given $m \times n$ matrix A, $A^+ = U_1 \Sigma_1^{-1} V_1^H$ is the Moore-Penrose Pseudoinverse of A.

Theorem 6.1. *For $m \times n$ A and vector $b \in \mathbb{F}^m$, there is a unique $x \in \mathcal{N}(A)$ such that $x = (A^H A)^{-1} A^H b$. Also, given the compact form of the SVD, $A = U_1 \Sigma_1 V_1^H$, then $x = A^+ b$.*

Proof. To see the equality just show $A^H A x = A^H A A^+ b == A^H b$, using the SVD to reduce terms. For uniqueness, suppose that v also solves the equation, then it must be that $A^H A x = A^H A v = A^H b$, so $A^H A(x - v) = 0$. Therefor $x - v$ is in the null space of $A^H A$ and thus the null space of A (equality of both null spaces was shown in chapter 3). But since neither vector is in the null space, their difference cannot be in the null space unless it is zero, therefore $v = x$.

Theorem 6.2. *Schmidt, Mirsky, Eckart-Young: Let A be an $m \times n$ matrix with rank r. Then*

$$\min_{\text{rank} B = s} ||A - B||^2 = \sigma_{s+1} \tag{8}$$

for any $s < r$. And argmin$= \sum_{i=1}^{s} \sigma_i u_i v_i^H$

7 Exercises

2,4,6,8,13,15,16,18,20,24,25,27,28,31,32,33,36,38