A Short Proof of the Spectral Theorem

Daniel Son

1. Introduction

The goal of this paper is to present a short proof of the Spectral theorem. Basic understanding of matrix and bases are assumed.

2. Eigenvalues and Eigenvectors

Let A be a square matrix of order n where entries are complex numbers. It might be the case that for some nonzero column vector $\vec{x}$ and some complex scalar λ, the following holds:

$$A\vec{x} = \lambda\vec{x}$$

We call λ to be the **Eigenvalue** of A, and $\vec{x}$ the **Eigenvector** corresponding to the scalar λ. Notice that if $\vec{x}$ is the Eigenvector of λ, so is $k\vec{x}$ for some arbitrary scalar k.

Also, we establish the following proposition:

Proposition 2.1 Number of Eigenvalues

For square matrix A of order n, there are maximum n distinct Eigenvalues.

Proof. Write:

$$A\vec{x} = \lambda I\vec{x} \quad \text{and} \quad (A - \lambda I)\vec{x} = 0$$

This equation holds only if the matrix $A - \lambda I$ is singular. In order for this to be possible, the determinant of $A - \lambda I$ must be zero. In fact, the determinant is some polynomial of degree n. By the fundamental theorem of algebra, there are n distinct roots for this polynomial. $\square$

Note that the polynomial obtained from the determinant is called the **characteristic polynomial** of matrix A.

To build up the spectral theorem, we are interested in a specific class of square matricies. Define the **spectrum** of matrix A to be the set of its Eigenvalues. Nice properties occur if the spectrum of the matrix includes exactly n distinct elements, where n is the order of A.

Proposition 2.2 Distinct Eigenvalues imply linear independence of Eigenvectors

Let A be a square matrix of order n which has a spectrum of n distinct eigenvalues. For each eigenvalue, choose one nonzero eigenvector to create a set of vectors. This set is linearly independant.

Proof. Assume for a contradiction and suppose the eigenvectors $\{x_1, ...x_n\}$ are linearly dependant and that the spectrum is distinct. By the definition of linear dependance, write:

$$c_1\vec{x_1} + c_2\vec{x_2} + ...c_n\vec{x_n} = 0$$

and without loss of generality, assume $c_n \neq 0$. Let λ_i denote the eigenvalue corresponding to the eigenvector $\vec{x_i}$. Multiply the whole equation, first by the matrix A and then by the scalar λ_1. Write:

$$c_1A\vec{x_1} + ...c_nA\vec{x_n} = 0 \quad \text{and} \quad c_1\lambda_1\vec{x_1} + ...c_n\lambda_n\vec{x_n} = 0$$

$$c_1\lambda_1\vec{x_1} + ...c_n\lambda_1\vec{x_n} = 0$$

Subtract the top equation from the bottom to obtain:

$$c_2(\lambda_2 - \lambda_1)\vec{x_2} + ...c_n(\lambda_n - \lambda_1)\vec{x_n} = 0$$

Applying a series of such manipulations, we reach:

$$c_n \prod_{i=1}^{n-1} (\lambda_{i+1} - \lambda_i)\vec{x_n} = 0$$

For we assumed $c_n \neq 0$, it must be the case that one of the terms of the product must be zero, implying that the spectrum is not distinct. We reach a contradiction. □

3. Diagonalization

We continue our discourse with matrix diagonalization. We first introduce the definition of diagonal matricies and a shorthand notation.

Definition 3.1 Diagonal Matricies

A square matrix is called to be diagonal if all of its entries other than the elements in its main diagonal (that is the entries where $r = c$) are zero. Also, we denote:

$$diag\{\lambda_1, ...\lambda_n\} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

We now introduce diagonalization

Definition 3.2 Diagonalization

Let A be a square matrix of order n. If there exists a non-singular matrix S and a diagonal matrix D such that:

$$A = SDS^{-1}$$

we call the expansion as a diagonalization of matrix A.

Such an expansion is not guaranteed to exist. Nonetheless, if one exists, we call matrix A to be **non-defective**.

In fact, the propertiy of eigenvectors guarantee a nice technique for diagonalization for matricies with distinct spectrum.

Theorem 3.3 Eigendecomposition

Let matrix A be a square matrix with distinct spectrum. A is non-defective and its decomposition is $S\Lambda S^{-1}$ where S is a matrix generated by stacking column vector representation of the eigenvectors.

Proof. By proposition 2.2, we establish that the n eigenvectors are linearly independant. Thus, by uniqueness of dimensions, the eigenvectors form a basis. Let $\bar{A}$ be the linear transformation corresponding to the matrix A. That means, A is the matrix of the transform $\bar{A}$ with respect to the standard basis of the space $\mathbb{C}^n$. Obtain Λ by taking the matrix of $\bar{A}$ with respect to the basis x_i, the eigenvectors. Note:

$$\Lambda = diag\{\lambda_1, ...\lambda_n\}$$

A simple base conversion will allow us to rewrite A in terms of a product involving Λ. Let S be the matrix that converts the base from the elementary base x_i to the eigenvector base e_i. In other words, S is the matrix of the base x_i with respect to the basis e_i. We deduce:

$$A = S\Lambda S^{-1} \quad \text{and} \quad S = [\vec{x_1}, ...\vec{x_n}]$$

where $\vec{x_i}$ is the column vector of x_i with respect to the standard basis. □

4. Orthogonal matricies We wish to obtain a simpler expression for the matrix S^{-1} presented above. To continue this endeavor, we consider inner product of vectors.

Definition 4.1 Inner Product of vectors

Let a, b be vectors in an n-dimensional space with elementary base $\{e_i\}$. If

$$a = \sum_{i=1}^{n} a_i e_i \quad \text{and} \quad b = \sum_{i=1}^{n} b_i e_i$$

then the inner product of the two vector is defined as:

$$a \cdot b = \sum_{i=1}^{n} a_i b_i$$

An ordered set of vectors $\{v_1, ...v_n\}$ is called to be **orthonormal** if:

$$v_i \cdot v_j = \delta_{ij}$$

Where δ_{ij} is the Kronecker delta defined to be 1 if $i = j$ and 0 otherwise. Orthonormal bases exhibit an interesting property. By merely unfolding the definitions introduced above, we recognize that:

$$\sum_{k=1}^{n} v_{ik} v_{jk} = \delta_{ij}$$

Consider the matrix:

$$V := [\vec{v_1} ... \vec{v_n}] \quad \text{and} \quad V^T$$

The summation above implies that the product of the sum of the two matricies equals the identity matrix. V is invertible. Hence:

$$VV^T = I \quad \text{and} \quad V^T = V^{-1}$$

We have established:

Theorem 4.2 The matrix of an orthonormal base has an inverse which is equal to the transpose of the matrix.

5. Proof of the Spectral Theorem We have now prepared all the machinery necessary to prove the theorem.

Theorem 5.1 Spectral Theorem Let A be a similar matrix, that is $A^T = A$, with a distinct spectrum. A is non-defective, and its diagonalization is given as:

$$A = S\Lambda S^T$$

Proof. Invoking theorem 3.3, it suffices to show that $S^{-1} = S^T$. By theorem 4.2, showing that the eigenbasis $\{x_i\}$ is orthonormal yields the desired result.

From part 1, recall that the eigenvector can be scaled by an arbitrary scalar. Hence, it is possible to normalize any eigenvector such that $x_i \cdot x_i = 1$ for all $i \leq n$. To show normality, consider two different eigenvectors x_i and x_j with eigenvalues $\lambda_i \neq \lambda_j$.

Let $\vec{x_i}$, $\vec{x_j}$ be the column vectors of the two eigenvectors with respect to the standard basis. Write:

$$(A\vec{x_i})^T\vec{x_j} = \lambda_i({\vec{x_i}}^T \cdot \vec{x_j}) \quad \text{and} \quad (A\vec{x_i})^T\vec{x_j} = {\vec{x_i}}^T A^T \vec{x_j} = {\vec{x_i}}^T A\vec{x_j}$$

Some more manipulation on the right equation shows:

$${\vec{x_i}}^T A\vec{x_j} = {\vec{x_i}}^T \lambda_j \vec{x_j} = \lambda_j({\vec{x_i}}^T \cdot \vec{x_j})$$

Using the associative property of matrix multiplication along with $A^T = A$, we write:

$$\lambda_i({\vec{x_i}}^T \cdot \vec{x_j}) = \lambda_j({\vec{x_i}}^T \cdot \vec{x_j}) \quad \text{or} \quad (\lambda_i - \lambda_j)({\vec{x_i}}^T \cdot \vec{x_j}) = 0$$

And the spectrum of A is distinct. Thus we arrive at:

$${\vec{x_i}}^T \cdot \vec{x_j} = 0$$

for $i \neq j$. This proves that the eigenbasis is orthonormal, which in turn proves the theorem. □