The square matrix theorem, eigenvectors, eigenvalues, and diagonalization

Throughout this note, A denotes an $n \times n$ matrix. We begin by recalling the Square Matrix Theorem without proof.

Theorem 1 (Square Matrix)**.** *For an $n \times n$ matrix A, the following are equivalent.*
(1) A is invertible.
(2) A is a product of elementary matrices.
(3) A is row-equivalent to I_n.
(4) $A\vec{x} = \vec{b}$ has a unique solution for every $\vec{b} \in \mathbb{R}^n$.
(5) $A\vec{x} = \vec{0}$ implies $\vec{x} = \vec{0}$.
(6) The columns of A are linearly independent.
(7) The columns of A form a basis of $\mathbb{R}^n$.
(8) $\text{nullity}(A) = 0$.
(9) $\text{rank}(A) = n$.

Note that conditions (5) and (6) are basically the same information, since

$$A\vec{x} = \left(\begin{array}{c|c|c} A_1 & \cdots & A_n \end{array} \right) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 A_1 + \cdots + x_n A_n.$$

Exercise 2. Write out a proof of the Square Matrix Theorem. Then find more equivalent statements and add them to the list above.

Definition 3. Recall that an *eigenvector* for A corresponding to the *eigenvalue* λ is a $\vec{v} \neq \vec{0}$ satisfying the *eigenvalue equation*:

$$A\vec{v} = \lambda\vec{v}.$$

If we have k eigenvectors $\vec{v}_1, \ldots, \vec{v}_k$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_k$ respectively, we can consider the k eigenvalue equations as a single equation called the *multiple eigenvalue equation*:

$$\begin{aligned} A \underbrace{\left(\begin{array}{c|c|c} \vec{v}_1 & \cdots & \vec{v}_k \end{array} \right)}_{S} &= \left(\begin{array}{c|c|c} A\vec{v}_1 & \cdots & A\vec{v}_k \end{array} \right) \\ &= \left(\begin{array}{c|c|c} \lambda_1\vec{v}_1 & \cdots & \lambda_k\vec{v}_k \end{array} \right) \\ &= \underbrace{\left(\begin{array}{c|c|c} \vec{v}_1 & \cdots & \vec{v}_k \end{array} \right)}_{S} \underbrace{\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix}}_{D}. \end{aligned}$$

That is, $AS = SD$ where

$$S := \left(\begin{array}{c|c|c} \vec{v}_1 & \cdots & \vec{v}_k \end{array} \right) \qquad \text{and} \qquad D := \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix}.$$

Often, we will add the stipulation that the columns of S should be linearly independent to avoid including redundant eigenvectors.

We now state the important theorem that follows immediately by analyzing the multiple eigenvalue equation.

Theorem 4. *For an $n \times n$ matrix A, the following are equivalent:*
(1) A is similar to a diagonal matrix D via the invertible matrix S.
(2) There is a basis of $\mathbb{R}^n$ consisting of eigenvectors for A.

Proof. Suppose $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is a maximal set of linearly independent eigenvectors for A, and let $\lambda_1, \ldots, \lambda_k$ be the corresponding eigenvalues. Form the $n \times k$ and $k \times k$ matrices

$$S := \left(\, \vec{v}_1 \, \middle| \, \cdots \, \middle| \, \vec{v}_k \, \right) \qquad \text{and} \qquad D := \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix}$$

so that $AS = SD$. By the Square Matrix Theorem 1, the following are equivalent:

- $k = n$,
- the eigenvectors $\{\vec{v}_1, \ldots, \vec{v}_k\}$ for A form a basis of $\mathbb{R}^n$, and
- S is invertible.

Hence there is a basis of $\mathbb{R}^n$ consisting of eigenvectors for A if and only if $S^{-1}AS = D$ with S and D defined as above. □

Let us now give one easy example of a matrix that is diagonalizable. We will use the following lemma.

Lemma 5. *Suppose $\vec{v}_1, \ldots, \vec{v}_k$ are eigenvectors for A corresponding to eigenvalues $\lambda_1, \ldots, \lambda_k$, so that $A\vec{v}_i = \lambda_i \vec{v}_i$ for all $i = 1, \ldots, k$. If the λ_i are all distinct, then $\vec{v}_1, \ldots, \vec{v}_k$ are linearly independent.*

Proof. We proceed by induction on k. If $k = 1$, the statement is trivial, since a single eigenvector $\vec{v} \neq \vec{0}$. Suppose the statement holds true for any collection of $k - 1 \in \mathbb{N}$ distinct eigenvalues and eigenvectors. Suppose $c_1, \ldots, c_k \in \mathbb{R}$ such that

$$\vec{0} = \sum_{i=1}^{k} c_i \vec{v}_i. \tag{1}$$

Apply A to both sides to see that

$$\vec{0} = A \sum_{i=1}^{k} c_i \vec{v}_i = \sum_{i=1}^{k} c_i A \vec{v}_i = \sum_{i=1}^{k} c_i \lambda_i \vec{v}_i. \tag{2}$$

We now multiply Equation (1) by λ_k and subtract it from Equation (2) to get

$$\vec{0} = \lambda_k \sum_{i=1}^{k} c_i \vec{v}_i - \sum_{i=1}^{k} c_i \lambda_i \vec{v}_i = \sum_{i=1}^{k} (\lambda_k - \lambda_i) c_i \vec{v}_i = \sum_{i=1}^{k-1} (\lambda_k - \lambda_i) c_i \vec{v}_i.$$

Notice we now have a linear combination of $k-1$ eigenvectors $\vec{v}_1, \ldots, \vec{v}_{k-1}$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_{k-1}$ which is equal to $\vec{0}$. By the induction hypothesis, we have that $(\lambda_k - \lambda_i)c_i = 0$ for all $1 \leq i \leq k-1$. Now since $\lambda_1, \ldots, \lambda_k$ are distinct, we have that $\lambda_k - \lambda_i \neq 0$ for all $1 \leq i \leq k-1$. This means we can divide by $\lambda_k - \lambda_i$ to get $c_i = 0$ for all $1 \leq i \leq k-1$. We now see that Equation (1) is the equation $c_k \vec{v}_k = 0$, and since $\vec{v}_k \neq \vec{0}$, we must also have $c_k = 0$. Hence $\vec{v}_1, \ldots, \vec{v}_k$ are linearly independent. By the Principle of Mathematical Induction, we have proved the lemma. □

Corollary 6. *Suppose A has n distinct eigenvalues. Then A is diagonalizable.*

Proof. By Lemma 5, any set of eigenvectors $v_1, \ldots, v_n$ corresponding to the n distinct eigenvalues of A will be linearly independent, and thus a basis of $\mathbb{R}^n$. By Theorem 4, A is diagonalizable. □

Definition 7. Let A be an $n \times n$ matrix, and recall the characteristic polynomial

$$p_A(\lambda) := \det(\lambda I - A) = \prod_{i=1}^{n} (\lambda - \lambda_i)$$

splits into linear factors by the Fundamental Theorem of Algebra. The λ_i are exactly the eigenvalues of A, counted with *algebraic* multiplicity. That is, if the distinct eigenvalues of A are $\lambda_1, \dots, \lambda_k$, we define the *algebraic multiplicity* of λ_i to be the $m_i \geq 1$ such that

$$p_A(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{m_i}.$$

For $1 \leq i \leq k$, the *eigenspace* associated to λ_i is

$$E_i := NS(A - \lambda_i I),$$

and we define the *geometric multiplicity* to be

$$g_i := \dim(E_i) = \dim(NS(A - \lambda_i I)).$$

Example 8. Consider the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Then $p_A(\lambda) = \lambda^2$, so A has one eigenvalue, namely zero. Its algebraic multiplicity is 2, while its geometric multiplicity is 1.

In more generality, consider the $n \times n$ matrix

$$A = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{pmatrix}$$

The matrix A has characteristic polynimial $p_A(\lambda) = \lambda^n$, so A has only zero as an eigenvalue. Its algebraic multiplicity is n, but its geometric multiplicity is 1.

We state the following theorem without proof.

Theorem 9. *Let A be an $n \times n$ matrix with characteristic polynomial*

$$p_A(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{m_i}.$$

The following are equivalent.

(1) The algebraic multiplicity m_i equals the geometric multiplicity g_i for all $1 \leq i \leq k$.

(2) A is diagonalizable.

Example 10. The matrices in Example 8 above are not diagonalizable.