Characterization of Diagonalizability via Minimal Polynomial

S. Kumaresan
School of Math. & Stat.
University of Hyderabad
Hyderabad 500 046
kumaresa@gmail.com

The aim of this article is to indicate direct and simple proofs of the following well-known and useful result.

Theorem 1. *Let V be an n-dimensional vector space over $\mathbb{F}$ and $T\colon V \to V$ be a linear map. Then T is diagonalizable iff the minimal polynomial of T splits into distinct linear factors in $\mathbb{F}[X]$.* □

We shall quickly review the technical words used in the statement and give proofs.

We say that a linear map $T\colon V \to V$ is a *diagonalizable* if there exist scalars $\alpha_k \in \mathbb{F}$, $1 \le k \le n$ and a basis $\{v_1, \dots, v_n\}$ of V such that $Tv_k = \alpha_k v_k$. Such a basis is called a T-eigen-basis of V or simply an eigen-basis of V.

Given a polynomial $f(X) \in \mathbb{F}[X]$, we define $f(T)$ by

$$f(T) := c_0 I + c_1 T + c_2 T^2 + \cdots + c_k T^k, \text{ if } f(X) = \sum_{j=0}^{k} c_j X^j.$$

Since $T \in L(V,V)$, an n^2 dimensional vector space, the set $\{T^k : 0 \le k \le n^2\}$ is linearly dependent in $L(V,V)$. Hence there exist scalar $c_i \in \mathbb{F}$, $0 \le i \le n^2$ such that $\sum_{j=0}^{n^2} c_j T^j = 0$, as linear maps. This says that there exists a polynomial $f(X) := \sum_{j=0}^{n^2} c_j X^j$ such that $f(T) = 0$. Thus the set of polynomials

$$I := \{p(X) \in \mathbb{F}[X] : p(T) = 0\}$$

is non-empty and it is easy to see that I is an ideal in $\mathbb{F}[X]$. Since $\mathbb{F}[X]$ is a PID, there exists a polynomial $\mu_T(X) \in I$ of least degree such that I is generated by $\mu_T(X)$. Let $k := \deg \mu_T$. If c_k is the (necessarily nonzero) coefficient of X^k, then $m_T(X) := c_k^{-1} \mu_T(X)$ is also a generator of I. Note that $m_T(X)$ is the unique polynomial in I with the properties: (i) $\deg m_T(X) \le \deg p(X)$ for any $p(X) \in I$ and that (ii) the leading coefficient (that is, the coefficient of the top degree term) is 1. The polynomial $m_T(X)$ is called the *minimal polynomial* of T.

We say that a polynomial $f(X) = \sum_{k=0}^{n} c_k X^k \in \mathbb{F}[X]$ *splits into distinct linear factors* if there exist n distinct elements $\alpha_j \in \mathbb{F}$, $1 \leq j \leq n$, and a $c \in \mathbb{F}$ such that $f(X) = c(X-\alpha_1)\cdots(X-\alpha_n)$. If $f(X)$ is monic then necessarily $c = 1$. With these definitions, the statement should make complete sense.

We also need a fact which readers must have used in integration of rational functions. We need only a very special case which we recall. Let $f(X) = (X-\alpha_1)\cdots(X-\alpha_n)$ be a polynomial with all the roots α_j distinct. We then can find constants $c_j \in \mathbb{F}$ such that

$$\frac{1}{f(X)} = \sum_{j=1}^{n} \frac{c_j}{X-\alpha_j}, \text{ in fact, we have } c_j = \frac{1}{f'(\alpha_j)}. \tag{1}$$

In case you are worried about the meaning of $f'(\alpha_j)$, you may take $\mathbb{F} = \mathbb{C}$. As a confidence building exercise, you may show that for $\alpha \neq \beta$

$$\frac{1}{(X-\alpha)(X-\beta)} = \frac{1}{\alpha-\beta}\left(\frac{1}{X-\alpha} - \frac{1}{X-\beta}\right).$$

If A is an $n \times n$ matrix,we know that we can view it as a linear map T_A on the vector space of column vectors $\mathbb{F}^n$. We define the minimal polynomial of A to be that of T_A.

A couple of easy exercises are in order.

Ex. 2. The minimal polynomial of a diagonal linear map $T \in L(V,V)$ is $(X-\alpha_1)\cdots(X-\alpha_k)$, where $\alpha_1, \ldots, \alpha_k$ are the distinct distinct eigenvalues of T.

Ex. 3. We say that two linear maps $A, B \in L(V,V)$ are similar if there exists an invertible $C \in L(V,V)$ such that $CAC^{-1} = B$. Then $m_A(X) = m_B(X)$. Analogous statement is true if we consider similar matrices.

We are now ready to prove the theorem.

Proof. Let $A \in L(V,V)$ be diagonalizable with distinct eigenvalues, say $\alpha_1, \ldots, \alpha_k$. Then, by Ex. 2, the minimal polynomial of T is $m_T(X) = (X-\alpha_1)\cdots(X-\alpha_k)$. Thus, $m_T(X)$ is a product of distinct linear factors.

Assume now that $m_T(X)$ is a product of distinct linear factors, say, $f(X) \equiv m_T(X) = (X-\alpha_1)\cdots(X-\alpha_k)$.

> **Strategy:** Let $f_j(X) = (X-\alpha_1)\cdots(X-\alpha_{j-1})(X-\alpha_{j+1})\cdots(X-\alpha_k)$, that is the factor $(X-\alpha_j)$ is left out in the expression of $f(X)$. Let $W_j := f_j(T)(V)$. Observe that $(T-\alpha_j)$ takes all of W_j to 0 so that T is is the scalar operator $\alpha_j I$ on W_j. We show that V is the direct sum of W_j's.

By (1), there exist constants $c_j \in \mathbb{F}$, $1 \leq j \leq k$, such that $\frac{1}{f} = \sum_{j=1}^{k} \frac{c_j}{X-\alpha_j}$. Hence we obtain

$$1 = c_1 f_1 + \cdots + c_k f_k. \tag{2}$$

It follows that $I = c_1 f_1(T) + \cdots + c_k f_k(T)$. Therefore, for any $v \in V$, we have $v = c_1 f_1(T)v + \cdots + c_k f_k(T)v$. If we let $W_j := f_j(T)(V)$, then each $v_j := f_j(T)v \in W_j$. Thus, V is a sum of the W_j's. Since $f(T) = (T - \alpha_j I) f_j(T) = 0$, observe that T is is the scalar operator $\alpha_j I$ on W_j.

We now claim the sum $V = W_1 + \cdots + W_k$ is direct. Let $0 = v_1 + \cdots + v_k \in W_1 + \cdots + W_k$. We have already observed that each of the v_j is an eigenvector, $Tv_j = \alpha_j v_j$. We know that any set of nonzero eigenvectors corresponding to distinct eigenvalues is linearly independent. (See Remark 4 below.) Hence we conclude that each $v_j = 0$.

We have thus concluded that V is the direct sum of $W_j \subset \ker(T - \alpha_j I)$. If we let B_j as a basis of W_j, then the union $B = \cup_{j=1}^k B_j$ is a basis of V and this is an eigen-basis by construction. That is, T is diagonalizable. □

Remark 4. Let λ_j, $1 \le j \le k$ be distinct eigenvalues of $T\colon V \to V$. Let v_j be nonzero vectors such that $Tv_j = \lambda_j v_j$, $1 \le j \le k$. Let $c_1 v_1 + \cdots + c_k v_k = 0$. Let $(T - \lambda_2 I) \cdots (T - \lambda_k I)$ act on both sides of this equation. We then arrive at $c_1(\lambda_1 - \lambda_2) \cdots (\lambda_1 - \lambda_k) v_1 = 0$. It follows that $c_1 = 0$. Similarly, other $c_j' s$ are zero.

Remark 5. If you are hard-core formalist, you may find the proof unsatisfactory! For, to prove (2), we made use of (1). The existence of partial fractions is almost never established in standard courses. It is done only empirically. Note that (2) is a Bezout type identity.

Since the set $\{f_1, \dots, f_k\}$ is coprime, (that is, the only common divisors are the units), the ideal generated by them in $\mathbb{F}[X]$ is the unit ideal $\mathbb{F}[X]$. Hence there exist $g_j \in \mathbb{F}[X]$ such that $g_1(X) f_1(X) + \cdots + g_k(X) f_k(X) = 1$. We still define $W_j := f_j(T)(V)$ and it is killed by $T - \alpha_j I$ as earlier. Now that we have proved (2) rather satisfactorily, the rest of the proof goes through as earlier.

Remark 6. How do you find the minimal polynomial of T? There are several ways to do it. You can read it off if you know the Jordan canonical form of T. Another method runs as follows. Fix a basis $\{v_1, \dots, v_n\}$ of V. Since $\{v_j, Tv_j, \dots, T^n v_j\}$ is linearly dependent, there exists a monic polynomial of smallest degree, say, $m_j(X)$ such that $m_j(T)(v_j) = 0$. The minimal polynomial $m_T(X)$ is the least common multiple of m_j's. You are asked to justify this and use it to find out the minimal polynomials of some matrices of small size.

Remark 7. If the matrix of T is the Jordan matrix $J_k(\lambda)$, then $m_T(X) = (X - \lambda)^k$. Assume that we know the Jordan matrix of T. Then we can read off the minimal polynomial of T. In fact, we have the following result.

The characteristic and minimal polynomials of T are

$$f_T(X) = \prod_{i=1}^n (X - \lambda_i)^{m_i} \text{ and } m_T(X) = \prod_{i=1}^n (X - \lambda_i)^{k_i}$$

where the m_i is the sum of the sizes (number of rows/columns) of the Jordan blocks with eigenvalue λ_i and k_i is the maximum size of such blocks.