Linear Algebra II - Annihilator of A Linear Transformation

Ming-Hsuan Kang

March 11, 2020

Recall that for a linear transformation $T : V \mapsto V$, its annihilator is

$$\text{Ann}(T) = \left\{ f(x) \in F[x] \middle| f(T) \equiv 0 \right\}$$

which always contains the characteristic polynomial of T by the Cayley-Hamilton Theorem. This set has some nice properties which are very useful for studying linear transforms.

First, let us recall some properties of polynomials, which you should already know in high school. (They will be proved again in the course of Abstract Algebra.) For two non-zero polynomials $f(x)$ and $g(x)$,

1. $h(x)$ is called a common divisor of $f(x)$ and $g(x)$ if $h(x)|f(x)$ and $h(x)|g(x)$;
2. there exists a unique monic (,i.e. its leading coefficient equals to one) greatest common divisor (GCD) $d(x)$ such that for all common divisor $h(x)$, we have $\deg h(x) \leq \deg d(x)$;
3. there exist two polynomials $a(x)$ and $b(x)$ such that $a(x)f(x) + b(x)g(x) = d(x)$. Especially, when $d(x) = 1$, $f(x)$ and $g(x)$ are called coprime.

Now for $f(T) \in \text{Ann}(T)$, suppose $f(x) = f_1(x)f_2(x)$ such that $f_1(x)$ and $f_2(x)$ are coprime. Then there exist two polynomials $a_1(x)$ and $a_2(x)$ such that $a_1(x)f_2(x) + a_2(x)f_1(x) = 1$. Let $P_1 = a_1(T)f_2(T)$ and $P_2 = a_2(T)f_1(T)$, then we have

$$P_1P_2 = a_2(T)f_1(T)a_1(T)f_2(T) \equiv 0 \qquad \text{and} \qquad P_1 + P_2 = I.$$

In this case, we have the following result.

Theorem 1. *Let $V_1 = \ker(f_1(T))$ and $V_2 = \ker(f_2(T))$.*

1. *V_1 and V_2 are T-invariant.*
2. *$V = V_1 \oplus V_2$.*
3. *P_1 is a projection onto V_1 along V_2.*
4. *P_2 is a projection onto V_2 along V_1.*

Proof. The part (1) is left to readers as exercise. For $v \in V$, let $\vec{v}_1 = P_1(\vec{v})$ and $\vec{v}_2 = P_2(\vec{v})$. First, we show that $\vec{v}_1 \in V_1$ and $\vec{v}_2 \in V_2$. Note that

$$f_1(T)(\vec{v}_1) = f_1(T)(P_1(\vec{v})) = f_1(T)\Big(a_1(T)f_2(T)(\vec{v})\Big) = a_1(T)\Big(f_1(T)f_2(T)\Big)(\vec{v}) = \vec{0}$$

which means $\vec{v}_1 \in V_1$. Similarly, we also have $\vec{v}_2 \in V_2$. Suppose $\vec{v} \in V_1 \cap V_2$, which means $f_1(T)(\vec{v}) = f_2(T)(\vec{v}) = \vec{0}$. Then

$$\vec{v} = (P_1 + P_2)(\vec{v}) = \Big(a_1(T)f_2(T) + a_2(T)f)_1(T)\Big)(\vec{v}) = a_1(T)f_2(T)(\vec{v}) + a_2(T)f_1(T)(\vec{v}) = \vec{0}.$$

Combining the above two results, we conclude that $V = V_1 \oplus V_2$.

We have shown that $\vec{v} = \vec{v}_1 + \vec{v}_2$ where $\vec{v}_1 = P_1(\vec{v}) \in V_1$ and $\vec{v}_2 = P_2(\vec{v}) \in V_2$. By the definition of projections, (3) and (4) hold. □

Corollary 2. *We have* $\ker(f_2(T)) = \mathrm{Im}(f_1(T))$ *and* $V = \ker(f_1(T)) \oplus \mathrm{Im}(f_1(T))$.

Proof. Note that we have

$$\mathrm{Im}(f_1(T)) \supseteq \mathrm{Im}\big(f_1(T)a_2(T)\big) = \mathrm{Im}(P_2) = V_2 = \ker(f_2(T)).$$

By the dimension formula and the part (2) in the above theorem, we have

$$\dim \mathrm{Im}(f_1(T)) = \dim V - \dim \ker(f_1(T)) = \dim \ker(f_2(T)).$$

We conclude that $\mathrm{Im}(f_1(T)) = \ker(f_2(T))$ and the part (2) in the above theorem can be rewritten as $V = \ker(f_1(T)) \oplus \mathrm{Im}(f_1(T))$. □

0.1 Genearalized kernel and generalized image

Consider the case $f(x) = f_T(x)$. Let $f_1(x) = x^m$ where m is the multiplicity of the zero eigenvalue (which can be equal to zero). By Corollary 2, we have the following nice decomposition:

$$V = \ker(T^m) \oplus \mathrm{Im}(T^m).$$

Remark. If $m = 0$, then $T^0 = I$, $\ker(T^m) = \{\vec{0}\}$, and $\mathrm{Im}(T^m) = V$.
Remark. In general, $\dim V = \dim \ker(T) + \dim \mathrm{Im}(T)$ always holds but $\ker(T) + \mathrm{Im}(T)$ may not be a direct sum.

Theorem 3. *Under the above condition, we have*

1. $\ker(T^m) = \bigcup_{i=1}^{\infty} \ker(T^i)$.
2. $\mathrm{Im}(T^m) = \bigcap_{i=1}^{\infty} \mathrm{Im}(T^i)$.

Proof. It suffices to show that $\ker(T^i) \subseteq \ker(T^m)$ and $\mathrm{Im}(T^i) \supseteq \mathrm{Im}(T^m)$ for all i. If $i \leq m$, then it is obvious. Suppose $i > m$. Since $f(x) = x^m f_2(x) \in \mathrm{Ann}(T)$, we also have $x^i f_2(x) \in \mathrm{Ann}(T)$. Applying Corollary 2 to both polynomials $x^m f_2(x)$ and $x^i f_2(x)$, we obtain

$$\ker(T^i) = \mathrm{Im}(f_2(T)) = \ker(T^m) \quad \text{and} \quad \mathrm{Im}(T^i) = \ker(f_2(T)) = \mathrm{Im}(T^m).$$

□

Definition The subspace $\bigcup_{i=1}^{\infty} \ker(T^i)$ is called the generalized kernel of T, denoted by $\ker_\infty(T)$ and $\bigcap_{i=1}^{\infty} \mathrm{Im}(T^i)$ is called the generalized image of T, denoted by $\mathrm{Im}_\infty(T)$.

Finally, let us summarize the above discussion as the following theorem.

Theorem 4. *The following holds.*

$$V = \ker_\infty(T) \oplus \mathrm{Im}_\infty(T).$$

Generalized Eigenspaces Decomposition

Suppose $f(T) \equiv 0$ and $f(x) = f_1(x) \cdots f_k(x)$ such that $f_i(x)$ and $f_j(x)$ are coprime for all $i \neq j$. For $i = 1$ to k, set

$$g_i(x) = f(x)/f_i(x).$$

One can show that $g_1(x), \cdots, g_2(x)$ are coprime, so there exist $a_1(x), \cdots, a_k(x)$ such that

$$a_1(x)g_1(x) + \cdots + a_k g_k(x) = 1.$$

Set $P_i = a_i(T)g_i(T)$, then we still have

$$P_1 \cdots P_k = 0 \qquad \text{and} \qquad P_1 + \cdots + P_k = I.$$

Theorem 5. *Let $V_i = \ker(f_i(T))$.*

1. *$V = V_1 \oplus \cdots \oplus V_k$.*
2. *P_i is a projection onto V_i along $V_1 \oplus \cdots \oplus \hat{V}_i \oplus \cdots \oplus V_k$.*

Proof. Exercise. □

Applying the above theorem, we obtain the following theorem immediately.

Corollary 6. *If there exists a split polynomial $f(x)$ without repeated roots in* $\mathrm{Ann}(T)$, *(In this case, $f(x) = \prod_{i=1}^{n}(x - \lambda_i)$ for some $\lambda_i \in F$.) Then T is diagonalizable.*

Proof. Exercise. □

Now suppose $f(x) = f_T(x) = \prod_{i=1}^{k}(x - \lambda_i)^{m_i}$ splits over F where $\lambda_1, \cdots, \lambda_k$ are all distinct. Set $f_i(x) = (x - \lambda_i)^{m_i}$. Then $f_i(x)$ and $f_j(x)$ are coprime for all $i \neq j$.Together with Theorem 3, Theorem 5, we have

$$V = \bigoplus_{i=1}^{k} \ker(T - \lambda_i I)^{m_i} = \bigoplus_{i=1}^{k} \ker_\infty(T - \lambda_i I).$$

Here

$$\ker_\infty(T - \lambda_i I) = \bigcup_{m=1}^{\infty} \ker(T - \lambda_i I)^m$$

is called the generalized eigenspace of T corresponding to λ_i, also denoted by $E_\infty(\lambda_i)$.
Remark. In particle, one shall use $E_\infty(\lambda_i) = \ker\left((T - \lambda_i I)^{m_i}\right)$ to compute $E_\infty(\lambda_i)$, where m_i is the algebraic multiplicity of λ_i.

We summarize the above discussion as the following theorem.

Theorem 7. *For a linear transform $T : V \mapsto V$, if its characteristic polynomial $f_T(x)$ splits, then V can be decomposed as the direct sum of generalized eigenspaces of T.*

Example Let $A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 2 & 1 & -1 \\ 1 & -2 & 4 & -1 \\ 2 & -1 & 0 & 3 \end{pmatrix}$. Then $f_A(x) = (x-2)^2(x-3)^2$.

For $\lambda = 2$,

$$E_\infty(2) = \ker(A-2I)^2 = \ker \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & -1 \\ 1 & -2 & 2 & -1 \\ 2 & -1 & 0 & 1 \end{pmatrix} = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

For $\lambda = 3$,

$$E_\infty(3) = \ker(A-3I)^2 = \ker \begin{pmatrix} -2 & 1 & 0 & 0 \\ -1 & -1 & 1 & -1 \\ 1 & -2 & 1 & -1 \\ 2 & -1 & 0 & 0 \end{pmatrix} = \operatorname{span}\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix} \right\}.$$

For the above example, we see the dimension of $E_\infty(\lambda_i)$ is equal to the algebraic multiplicity of λ_i. This is indeed a general fact.

Lemma 8. *The characteristic polynomial of T restricted on $E_\infty(\lambda_i)$ is equal to $(x-\lambda_i)^{d_i}$ where $d_i = \dim E_\infty(\lambda_i)$.*

Proof. Exercise. $\square$

Theorem 9. *Let λ be an eigenvalue of T of algebraic multiplicity m. Then $\dim E_\infty(\lambda)$ is equal to m. (Note that we do not need assume that $f_T(x)$ splits.)*

Proof. Write $f_T(x) = (x-\lambda)^m g(x)$ for some $g(x)$ with $g(\lambda) \neq 0$. Let $W_1 = E_\infty(\lambda)$ and $W_2 = \ker(g(T))$ for short. Then we have

$$V = \ker(T-\lambda)^m \oplus \ker(g(T))) = W_1 \oplus W_2.$$

From this decomposition together with the previous lemma, we obtain

$$f_T(x) = f_{T\big|_{W_1}}(x) f_{T\big|_{W_2}}(x) = (x-\lambda)^d f_{T\big|_{W_1}}(x)$$

where $d = \dim E_\infty(\lambda)$. Comparing with the factorization $(x-\lambda)^m g(x)$, we obtain that $d \leq m$ and if $d < m$, then $f_{T\big|_{W_2}}(\lambda)$ =0. Now suppose $d < m$, then λ is an eigenvalue of $T\big|_{W_2}$.

On the other hand, since $W_2 = \ker(g(T))$, $g(T\big|_{W_2}) = g(T)\big|_{W_2} \equiv 0$. Therefore, $g(x) \in \operatorname{Ann}(T\big|_{W_2})$, which implies that $g(\lambda) = 0$. This contradicts to the assumption of $g(x)$.

$\square$

0.2 Minimal Polynomial

Let us study which kinds of polynomials may occur in $\mathrm{Ann}(T)$.

Theorem 10. *Among nonzero polynomials in* $\mathrm{Ann}(T)$*, there exists a unique monic polynomial* $m_T(x)$ *in* $\mathrm{Ann}(T)$ *such that* $m_T(x)|f(x)$ *for all* $f(x) \in \mathrm{Ann}(T)$*.*

Proof. Let $m_T(x)$ be a (nonzero) monic polynomial in $\mathrm{Ann}(T)$ of minimal degree. First, let us show that such $m_T(x)$ is unique. Suppose there is another monic polynoial $f(x)$ in $\mathrm{Ann}(T)$ of the same degree, then one can obtain a monic polynomial with smaller degree in $\mathrm{Ann}(T)$ by rescaling $f(x) - m_T(x)$, which is a contradiction. Next, for a nonzero $f(x)$ in $\mathrm{Ann}(T)$, write $f(x) = q(x)m_T(x) + r(x)$ by long division algorithm. If $r(x) \neq 0$, then the degree of $r(x)$ is smaller than the degree of $m_T(x)$. On the other hand,

$$r(T) = f(T) - q(T)m_T(T) \equiv 0.$$

Thus $r(x) \in \mathrm{Ann}(T)$, which is a contradiction. □

The following theorem shows that the minimal polynomial can not be too "small".

Theorem 11. *Let* λ *be a zero of* $f_T(x)$ *over* $\mathbb{C}$*. For any* $f(x) \in \mathrm{Ann}(T)$*, we always have* $f(\lambda) = 0$*. Especially,* $m_T(\lambda) = 0$*.*

Proof. Let $A \in M_n(F)$ be a matrix representation of T. Then we have $f_T(x) = f_A(x)$ and $f(A) \equiv 0$. Regard A as a complex matrix and let $\vec{v}$ be the complex eigenvector corresponding to λ in $\mathbb{C}^n$. Write $f(x) = \sum a_i x^i$. Then

$$\vec{0} = f(A)\vec{v} = \sum a_i A^i \vec{v} = \sum a_i \lambda^i \vec{v} = f(\lambda)\vec{v}.$$

We conclude that $f(\lambda) = 0$. □

Immediately, we have the following two results.

Corollary 12. *Let* $\lambda_1, \cdots, \lambda_k$ *be all distinct zeros of* $f_T(x)$ *over* $\mathbb{C}$*. Then* $\prod_{i=1}^k (x - \lambda_i)$ *divides* $m_T(x)$*.*

Corollary 13. *Suppose* $f_T(x)$ *splits and* $\lambda_1, \cdots, \lambda_k$ *be all distinct zero of* $f_T(x)$*. Let* $f(x) = \prod_{i=1}^k (x - \lambda_i)$*. Then* T *is diagonalizable if and only if* $f(T) \equiv 0$*, or equivalently* $f(x) = m_T(x)$*.*

Finally, let us give an explicit way to find the minimal polynomial. Suppose $f_T(x) = \prod (x - \lambda_i)^{m_i}$. Then we have shown that the minimal polynomial is of the form $m(x) = \prod (x - \lambda_i)^{r_i}$, so the goal is to find the minimal r_i for each i so that $m(T) \equiv 0$.

If $m(T) \equiv 0$, then we have $\ker(T - \lambda_i I)^{r_i} = E_\infty(\lambda_i)$. (See the subsection of the generalized eigenspace.) On the other hand, if $\ker(T - \lambda_i I)^{r_i} = E_\infty(\lambda_i)$ for all i,

$$\ker(m(T)) \supseteq \ker(T - \lambda_i I)^{r_i} = E_\infty(\lambda_i) \quad \text{for all} i.$$

Therefore, $\ker(m(T)) \supseteq \oplus E_\infty(\lambda_i) = V$, which means $m(T) \equiv 0$. We conclude that

$$\begin{aligned} r_i &= \min\{m \in N | \ker(T - \lambda_i I)^m = E_\infty(\lambda_i)\} \\ &= \min\{m \in N | \dim \ker(T - \lambda_i I)^m = m_i\} \\ &= \min\{m \in N | \mathrm{rank}(T - \lambda_i I)^m = n - m_i\}. \end{aligned}$$

(Here m_i is the algebraic multiplicity of λ_i of T.)

Example Let $A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 2 & 1 & -1 \\ 1 & -2 & 4 & -1 \\ 2 & -1 & 0 & 3 \end{pmatrix}$. Then $f_A(x) = (x-2)^2(x-3)^2$.
Let $m_A(x) = (x-2)^{r_1}(x-2)^{r_2}$. For $\lambda = 2$,

$$\mathrm{rank}(A - 2I) = \mathrm{rank} \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & -1 \\ 1 & -2 & 2 & -1 \\ 2 & -1 & 0 & 1 \end{pmatrix} = 3$$

and

$$\mathrm{rank}(A - 2I)^2 = \mathrm{rank} \begin{pmatrix} 0 & -1 & 1 & -1 \\ 0 & -2 & 2 & -2 \\ 1 & -2 & 2 & -1 \\ 1 & 1 & -1 & 2 \end{pmatrix} = 2.$$

Thus, $r_1 = 2$.
Remark. There is another way to see $r_1 = 2$: From $\mathrm{rank}(A-2I) = 3$ and $m_2 = 2$, we know $r_1 > 1$ and $r_1 \leq 2$ which implies that $r_1 = 2$.
For $\lambda = 3$,

$$\mathrm{rank}(A - 3I) = \mathrm{rank} \begin{pmatrix} -2 & 1 & 0 & 0 \\ -1 & -1 & 1 & -1 \\ 1 & -2 & 1 & -1 \\ 2 & -1 & 0 & 0 \end{pmatrix} = 2.$$

Thus, $r_2 = 1$. We conclude that $m_A(x) = (x-2)^2(x-3)$.