

18.700 LECTURE NOTES, 11/12/04

CONTENTS

1.	Resultants	1
2.	Discriminants	3
3.	A fact about polynomials	5
4.	The fundamental theorem of algebra	6
5.	The regular matrices are dense	7
6.	The Cayley-Hamilton Theorem	8

1. RESULTANTS

Let \mathbb{F} be a field, let $d, e \geq 0$ be integers, and let $f(x), g(x) \in \mathbb{F}[x]$ be polynomials of degrees $\leq d$ and $\leq e$ respectively (possibly the zero polynomial). How can one tell if f and g have a common factor of positive degree?

Proposition 1.1. *Assume either f is nonzero of degree d or g is nonzero of degree e . There is a common factor of f and g of positive degree iff there exist polynomials f_1 and g_1 of degrees $< e$ and $< d$ respectively, not both zero, such that $f \cdot g_1 + g \cdot f_1 = 0$.*

Proof. (\Rightarrow) Suppose there exists a common factor of positive degree, say h . Denote $c = \deg(h) > 0$. Let $f_1 = (f/h)$ and $g_1 = -(g/h)$. Then f_1 and g_1 have degrees $< d$ and $< e$ respectively, and,

$$f \cdot g_1 + g \cdot f_1 = -x^{c-1}(fg/h) + x^{c-1}(fg/h) = 0.$$

Since at least one of f and g is nonzero, at least one of f_1 and g_1 is nonzero.

(\Leftarrow) Let f_1 and g_1 be polynomials of degrees $< d$ and $< e$ respectively such that $f \cdot g_1 + g \cdot f_1 = 0$, i.e., $f \cdot g_1 = -g \cdot f_1$, and which are not both zero. At least one of f and g is nonzero of degree d , resp. e ; without loss of generality, assume f is nonzero of degree d . If g_1 is nonzero, then $f \cdot g_1$ is nonzero, which implies f_1 is nonzero. Since at least one of f_1, g_1 is nonzero, it follows that f_1 is nonzero.

Consider the irreducible decomposition of f , say $f = p_1^{d_1} \cdots p_r^{d_r}$ where p_1, \dots, p_r are non-proportional irreducible polynomials of positive degree and $d_1 \cdot \deg(p_1) + \cdots + d_r \cdot \deg(p_r) = d$. For every $i = 1, \dots, r$, let $c_i \geq 0$ be the greatest integer such that $p_i^{c_i}$ divides f_1 . Then $c_1 \cdot \deg(p_1) + \cdots + c_r \cdot \deg(p_r) \leq \deg(f_1) < d$. So there exists $1 \leq i \leq r$ such that $c_i < d_i$. Then p_i divides $(f/p_i^{c_i}) \cdot g_1 = -g \cdot (f_1/p_i^{c_i})$, but p_i does not divide $(f_1/p_i^{c_i})$. Therefore p_i divides g , i.e., p_i is a polynomial of positive degree that factors both f and g . \square

This suggests a linear algebra solution to this polynomial problem.

Definition 1.2. Let $f, g \in \mathbb{F}[x]$ be polynomials of degrees $\leq d$ and $\leq e$. Let $P^{d-1} \oplus P^{e-1}$ be the \mathbb{F} -vector space of pairs (f_1, g_1) of polynomials of degrees $< d$ and $< e$ respectively, where $(f_1, g_1) + (f_2, g_2) = (f_1 + f_2, g_1 + g_2)$ and $a \cdot (f_1, g_1) = (a \cdot f_1, a \cdot g_1)$. The associated linear transformation is $T_{(f,d),(g,e)} : P^{d-1} \oplus P^{e-1} \rightarrow P^{d+e-1}$, $T_{(f,d),(g,e)}(f_1, g_1) = f \cdot g_1 + g \cdot f_1$, where P^{d+e-1} is the \mathbb{F} -vector space of polynomials of degree $d + e - 1$.

Corollary 1.3. Assume either f is nonzero of degree d or g is nonzero of degree e . The polynomials f and g have a common factor of positive degree iff $T_{(f,d),(g,e)}$ has nullity ≥ 1 .

Proof. The kernel of $T_{(f,d),(g,e)}$ is the set of pairs (f_1, g_1) such that $f \cdot g_1 + g \cdot f_1 = 0$. By Proposition 1.1, there exists such a pair iff f and g have a common factor of positive degree. \square

Let \mathcal{B} denote the ordered basis of $P^{d-1} \oplus P^{e-1}$,

$$\mathcal{B} = ((1, 0), (x, 0), \dots, (x^{d-1}, 0), (0, 1), (0, x), \dots, (0, x^{e-1})).$$

Let \mathcal{C} denote the ordered basis of P^{d+e-1} , $\mathcal{C} = (1, x, \dots, x^{d+e-1})$. Let $f(x) = a_d x^d + \dots + a_1 x + a_0$ and let $g(x) = b_e x^e + \dots + b_1 x + b_0$. Denote by $A_{(f,d),(g,e)}$ the $(d+e) \times (d+e)$ matrix,

$$A_{(f,d),(g,e)}(i, j) = \begin{cases} b_{j-i}, & 1 \leq i \leq d, \text{ and } i \leq j \leq i+e, \\ a_{j+d-i}, & d+1 \leq i \leq d+e, \text{ and } i-d \leq j \leq i, \\ 0, & \text{otherwise} \end{cases}$$

In other words, $A_{(f,d),(g,e)}$ is the partitioned matrix,

$$A(f, d), (g, e) = \begin{pmatrix} B_{d,(g,e)} \\ B_{e,(f,d)} \end{pmatrix},$$

where $B_{d,(g,e)}$ is the $d \times (d+e)$ matrix,

$$B_{d,(g,e)} = \begin{pmatrix} b_0 & b_1 & \dots & b_{d-1} & b_d & \dots & b_e & 0 & \dots & 0 \\ 0 & b_0 & \dots & b_{d-2} & b_{d-1} & \dots & b_{e-1} & b_e & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_0 & b_1 & \dots & b_{e+1-d} & b_{e+2-d} & \dots & b_e \end{pmatrix},$$

and where $B_{e,(f,d)}$ is the $e \times (d+e)$ matrix,

$$B_{e,(f,d)} = \begin{pmatrix} a_0 & a_1 & \dots & a_d & 0 & \dots & 0 & \dots & 0 \\ 0 & a_0 & \dots & a_{d-1} & a_d & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & a_0 & \dots & a_d \end{pmatrix}.$$

Lemma 1.4. The matrix representative $[T_{(f,d),(g,e)}]_{\mathcal{C}, \mathcal{B}}$ is $A_{(f,d),(g,e)}$.

Proof. For every $i = 1, \dots, d$, $T_{(f,d),(g,e)}(x^{i-1}, 0) = x^{i-1}g$. The coordinate vector of $x^{i-1}g$ with respect to \mathcal{C} is the vector of coefficients. Thus $[x^{i-1}g]_j$ is the coefficient of x^{j-1} in $x^{i-1}g$, i.e., the coefficient of x^{j-i} in g . This is 0 unless $0 \leq j-i \leq e$, and then it equals b_{j-i} .

Similarly, for $i = d+1, \dots, d+e$, $T_{(f,d),(g,e)}(0, x^{i-d-1}) = x^{i-d-1}f$. The coordinate vector of $x^{i-d-1}f$ with respect to \mathcal{C} is the vector of coefficients. Thus $[x^{i-d-1}f]_j$

is the coefficient of x^{j-1} in $x^{i-d-1}f$, i.e., the coefficient of x^{j+d-i} in f . This is 0 unless $0 \leq j + d - i \leq d$, and then it equals a_{j+d-i} . \square

Definition 1.5. The *resultant* of (f, d) and (g, e) , $\text{Res}((f, d), (g, e))$ is the determinant of the $(d + e) \times (d + e)$ matrix $A_{(f,d),(g,e)}$.

Theorem 1.6. Assume either f is nonzero of degree d or g is nonzero of degree e . The polynomials f and g have a common factor iff $\text{Res}((f, d), (g, e)) = 0$.

Proof. By Corollary 1.3, f and g have a common factor iff $\text{null}(T_{(f,d),(g,e)}) > 0$. Of course, $\text{null}(T_{(f,d),(g,e)}) = \text{null}(A_{(f,d),(g,e)})$. Because $A_{(f,d),(g,e)}$ is a square matrix, by Theorem 2.1.12 it has a nonzero nullvector iff the determinant is 0, i.e., iff $\text{Res}((f, d), (g, e)) = 0$. \square

Example 1.7. Let $d = e = 2$. The resultant of $f(x) = a_2x^2 + a_1x + a_0$ and $g(x) = b_2x^2 + b_1x + b_0$ is,

$$\text{Res}((f, 2), (g, 2)) = \det \begin{pmatrix} b_0 & b_1 & b_2 & 0 \\ 0 & b_0 & b_1 & b_2 \\ a_0 & a_1 & a_2 & 0 \\ 0 & a_0 & a_1 & a_2 \end{pmatrix},$$

i.e.,

$$\text{Res}((f, 2), (g, 2)) = a_0^2b_2^2 - a_0a_1b_1b_2 + a_0a_2(b_1^2 - 2b_0b_2) + a_1^2b_0b_2 - a_1a_2b_0b_1 + a_2^2b_0^2.$$

Just as a further example, if $f = (x-1)(x+1) = x^2 - 1$ and $g = (x-1)^2 = x^2 - 2x + 1$, then $\text{Res}((f, 2), (g, 2))$ equals,

$$\begin{aligned} 1^2 \cdot 1^2 - 1 \cdot 0 \cdot (-2) \cdot 1 + 1 \cdot (-1)((-2)^2 - 2 \cdot 1 \cdot 1) + 0^2 \cdot 1 \cdot 1 - 0 \cdot (-1) \cdot 1 \cdot (-2) + (-1)^2 \cdot 1^2 \\ = 1 + 0 - 2 + 0 + 0 + 1 = 0. \end{aligned}$$

Also, if $f = (x-1)(x+1) = x^2 - 1$ and $g = x = 0x^2 + 1x + 0$, considered as a polynomial of degree ≤ 2 , then $\text{Res}((f, 2), (g, 2))$ equals,

$$\begin{aligned} 1^2 \dots 0 - 1 \cdot 0 \cdot 1 \cdot 0 + 1 \cdot (-1)(1^1 - 2 \cdot 0 \cdot 0) + 0^2 \cdot 0 \cdot 0 - 0 \cdot (-1) \cdot 0 \cdot 1 + (-1)^1 \cdot 0^1 \\ = 0 - 0 + (-1) + 0 - 0 + 0 = -1. \end{aligned}$$

This is nonzero, as dictated by Theorem 1.6.

2. DISCRIMINANTS

A particularly interesting case is when $g = f'$, the formal derivative of f with respect to x .

Definition 2.1. The *formal derivative with respect to x* is the unique \mathbb{F} -linear transformation $d/dx : \mathbb{F}[x] \rightarrow \mathbb{F}[x]$ such that $d/dx(x^n) = nx^{n-1}$ for every $n \geq 1$, and $d/dx(1) = 0$.

Remark 2.2. If $\mathbb{F} = \mathbb{R}$ or \mathbb{C} , the formal derivative can be interpreted as the usual derivative of the associated function, or the holomorphic derivative of the associated holomorphic function. But for \mathbb{F} a general field, e.g., $\mathbb{Z}/p\mathbb{Z}$, there is no such interpretation. In this case the formal derivative is simply the linear transformation defined above.

Lemma 2.3. The formal derivative with respect to x is the unique \mathbb{F} -linear transformation such that both,

- (i) $d/dx(x) = 1$, and,

(ii) for every $f(x), g(x) \in \mathbb{F}[x]$, $d/dx(f \cdot g) = f \cdot d/dx(g) + g \cdot d/dx(f)$.

Proof. First of all, $d/dx(x) = 1x^{1-1} = 1x^0 = 1$. By linearity, to prove that $d/dx(f \cdot g) = f \cdot d/dx(g) + g \cdot d/dx(f)$, it suffice to consider the case that $f = x^m$ and $g = x^n$. Up to relabeling, assume that $m \leq n$. If $m = 0$, then $f = 1$ and $d/dx(f \cdot g) = d/dx(g) = 1 \cdot d/dx(g)$. Because $d/dx(f) = 0$, this equation is $d/dx(f \cdot g) = f \cdot d/dx(g) + g \cdot d/dx(f)$.

Therefore assume $m \geq 1$. Then $d/dx(x^m \cdot x^n) = d/dx(x^{m+n})$, which by definition equals $(m+n)x^{m+n-1}$. Also $d/dx(f) = d/dx(x^m) = mx^{m-1}$ and $d/dx(g) = d/dx(x^n) = nx^{n-1}$. Therefore,

$$d/dx(f \cdot g) = (m+n)x^{m+n-1} = x^m \cdot (nx^{n-1}) + x^n \cdot (mx^{m-1}) = f \cdot d/dx(g) + g \cdot d/dx(f).$$

Next suppose that $T : \mathbb{F}[x] \rightarrow \mathbb{F}[x]$ is a linear transformation such that $T(x) = 1$ and $T(f \cdot g) = f \cdot T(g) + g \cdot T(f)$ for every $f, g \in \mathbb{F}[x]$. First of all,

$$T(1) = T(1 \cdot 1) = 1 \cdot T(1) + 1 \cdot T(1) = T(1) + T(1).$$

By cancellation, $T(1) = 0$. The claim is that for every integer $n \geq 0$, $T(x^n) = nx^{n-1}$. This will be proved by induction on n . For $n = 1$, this is the first hypothesis. Therefore assume $n > 1$ and the result is true for $n - 1$. Then, by the second hypothesis,

$$T(x^n) = T(x \cdot x^{n-1}) = x \cdot T(x^{n-1}) + x^{n-1} \cdot T(x).$$

By the induction hypothesis, $T(x^{n-1}) = (n-1)x^{n-2}$. By the first hypothesis, $T(x) = 1$. Thus,

$$T(x^n) = x \cdot (n-1)x^{n-2} + x^{n-1} \cdot 1 = n \cdot x^{n-1}.$$

Therefore the claim is proved by induction on n . Since $1, x, \dots$ is a basis for $\mathbb{F}[x]$ and $T(x^n) = d/dx(x^n)$ for every $n \geq 0$, the linear transformation T equals d/dx . \square

Proposition 2.4. *If $f(x)$, a polynomial of positive degree, factors as a product of linear factors, then $f(x)$ has a repeated linear factor iff $f(x)$ and $d/dx(f(x))$ have a common factor of degree $d \geq 1$.*

Proof. Let $f(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_s)^{e_s}$, where $\lambda_1, \dots, \lambda_s$ are distinct and e_1, \dots, e_s are positive integers. Then,

$$f'(x) = \sum_{i=1}^s e_i \cdot (x - \lambda_1)^{e_i-1} \cdot \prod_{1 \leq j \leq s, j \neq i} (x - \lambda_j)^{e_j}.$$

If $e_i > 1$, then each term in the sum is divisible by $(x - \lambda_i)^{e_i-1}$. So $(x - \lambda_i)^{e_i-1}$ is a common factor of $f(x)$ and $f'(x)$ of positive degree.

On the other hand, suppose each $e_i = 1$. Then for every $i = 1, \dots, s$, $x - \lambda_i$ divides,

$$g_i = \sum_{1 \leq j \leq s, j \neq i} \prod_{1 \leq k \leq s, k \neq j} (x - \lambda_k).$$

If $x - \lambda_i$ divides $f'(x)$, then $x - \lambda_i$ divides the difference,

$$f'(x) - g_i(x) = \prod_{1 \leq j \leq s, j \neq i} (x - \lambda_j),$$

which is ridiculous. Therefore no $x - \lambda_i$ divides $f'(x)$. Because the irreducible factors of $f(x)$ are $x - \lambda_1, \dots, x - \lambda_s$, there is no common factor of $f(x)$ and $f'(x)$ of positive degree. \square

Definition 2.5. Let $f(x)$ be a nonzero polynomial of degree $e \geq 1$. The *discriminant* of f is $\text{Disc}(f) = \text{Res}((f, e), (df/dx, e - 1))$.

Theorem 2.6. Suppose that $f(x)$ factors as a product of linear factors. Then $f(x)$ has a repeated linear factor iff $\text{Disc}(f) = 0$.

Proof. By Proposition 2.4, $f(x)$ has a repeated linear factor iff f and df/dx have a common factor of positive degree. The derivative df/dx is a polynomial of degree $\leq e - 1$. By Theorem 1.6, f and df/dx have a common factor of positive degree iff the resultant $\text{Res}((f, e), (df/dx, e - 1))$ is zero. \square

Example 2.7. Let $f(x) = ax^2 + bx + c$ where $a \neq 0$. Then $f'(x) = 2ax + b$ and the discriminant of f is $\text{Res}((f, 2), (f', 1))$,

$$\det \begin{pmatrix} c & b & a \\ b & 2a & 0 \\ 0 & b & 2a \end{pmatrix},$$

i.e.,

$$\text{Disc}(f) = c(4a^2) - b(2ba - ba) = a(4ac - b^2).$$

Because a is nonzero, it is customary to factor it from the expression, giving $(-1/a) \cdot \text{Disc}(f) = b^2 - 4ac$, familiar from the quadratic formula.

Example 2.8. Let $f(x)$ be a cubic polynomial. If 2 and 3 are different from 0 in \mathbb{F} , there is a straightforward linear change of coordinates $x^{\text{new}} = \mu x^{\text{old}} - \lambda$ so that $f(x)$ has the form,

$$f(x) = x^3 + ax + b.$$

The derivative of $f(x)$ is $f'(x) = 3x^2 + a$. So the discriminant of f is $\text{Res}((f, 3), (f', 2))$,

$$\det \begin{pmatrix} b & a & 0 & 1 & 0 \\ 0 & b & a & 0 & 1 \\ a & 0 & 3 & 0 & 0 \\ 0 & a & 0 & 3 & 0 \\ 0 & 0 & a & 0 & 3 \end{pmatrix},$$

i.e., expanding down the first column,

$$\text{Disc}(f) = 27b^2 + 4a^3.$$

In particular, if $f(x) = (x - 1)^2(x + 2) = x^3 - 3x + 2$, the discriminant is $27(2)^2 + 4(-3)^3 = 108 - 108 = 0$. And for $f(x) = (x - 1)(x + 1)x = x^3 - x$, the discriminant is $27(0)^2 + 4(-1)^3 = -4$. Observe that -4 equals 0 in \mathbb{F} iff $2 = 0$, i.e., $-1 = 1$, in which case $f(x) = (x - 1)^2x$ does have a repeated linear factor.

3. A FACT ABOUT POLYNOMIALS

Let x_1, \dots, x_n be variables, and let $f(x_1, \dots, x_n) \in \mathbb{F}[x_1, \dots, x_n]$ be a polynomial in n variables. There is an associated function $f : \mathbb{F}^n \rightarrow \mathbb{F}$ by substituting in for the variables.

Proposition 3.1. If \mathbb{F} has infinitely many elements, e.g., $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, then for every integer $n \geq 0$, the only polynomial in n variables that gives rise to the zero function is the zero polynomial.

Proof. For $n = 0$, $f = a$ for some $a \in \mathbb{F}$. The associated function sends the singleton set $\mathbb{F}^0 = \{*\}$ to $a \in \mathbb{F}$. Thus f is the zero function iff $a = 0$ iff f is the zero polynomial.

Next assume $n = 1$. Clearly the zero polynomial gives the zero function. Assume $f(x)$ is a nonzero polynomial, i.e., $f(x) = a_d x^d + \cdots + a_1 x + a_0$ where $d \geq 0$ and $a_d \neq 0$. Since $\deg(f) = d$, $f(x)$ has at most d distinct linear factors, so at most d distinct roots. But \mathbb{F} has infinitely many elements. Therefore there exists $\lambda \in \mathbb{F}$ such that $f(\lambda) \neq 0$.

By way of induction, assume $n > 1$ and assume the result is known for $n - 1$. Clearly the zero polynomial gives rise to the zero function. Thus assume f is not the zero polynomial. The polynomial f can be expanded as,

$$f(x_1, \dots, x_n) = \sum_{i=0}^d f_i(x_1, \dots, x_{n-1}) x_n^i,$$

where $d \geq 0$ and $f_d(x_1, \dots, x_{n-1})$ is a nonzero polynomial. By the induction hypothesis, there exist $\lambda_1, \dots, \lambda_{n-1} \in \mathbb{F}$ such that $f_d(\lambda_1, \dots, \lambda_{n-1}) \neq 0$. Consider the polynomial,

$$g(x) = f(\lambda_1, \dots, \lambda_{n-1}, x) = \sum_{i=0}^d a_i x^i.$$

By construction, $a_d \neq 0$. So by the same argument as in the last paragraph, there exists $\lambda \in \mathbb{F}$ such that $g(\lambda) \neq 0$. Therefore $f(\lambda_1, \dots, \lambda_{n-1}, \lambda) \neq 0$, which proves the proposition by induction on n . \square

Corollary 3.2. *Let $f(x) \in \mathbb{Z}[x_1, \dots, x_n]$ be a polynomial with integer coefficients. Then $f(x)$ is the zero polynomial iff the associated function $f : \mathbb{C}^n \rightarrow \mathbb{C}$ is the zero function.*

Proof. Thinking of integers as just special complex numbers, the polynomial f is also a polynomial in $\mathbb{C}[x_1, \dots, x_n]$. For every n -tuple of nonnegative integers, $e_0, \dots, e_n \geq 0$, the coefficient of $x_0^{e_0} \cdots x_n^{e_n}$ in f is 0 considered as an integer iff it is 0 considered as a complex number. Therefore f is 0 as an element in $\mathbb{Z}[x_1, \dots, x_n]$ iff f is 0 as an element in $\mathbb{C}[x_1, \dots, x_n]$. By Proposition 3.1, f is 0 as an element in $\mathbb{C}[x_1, \dots, x_n]$ iff f is 0 as a function. \square

4. THE FUNDAMENTAL THEOREM OF ALGEBRA

Definition 4.1. A field \mathbb{F} is *algebraically closed* if every nonconstant polynomial in $\mathbb{F}[x]$ factors as a product of linear polynomials.

Theorem 4.2 (The fundamental theorem of algebra). *The field of complex numbers is algebraically closed.*

This is an important theorem about the complex numbers. Every proof has some element of analysis. A simple proof, given in most complex analysis courses, is the following: Let $f(z)$ be a polynomial of degree $d \geq 1$ with no zero. Then $1/f(z)$ is a holomorphic function on \mathbb{C} . For every $\epsilon > 0$, there exists $R > \max(1, |a_0|/2d, \dots, |a_{d-1}|/2d, (2/\epsilon|a_d|)^{1/d})$ such that if $|z| \geq R$, then $|f(z)| = |a_d||z|^d|1 + a_{d-1}/z + \dots| \geq |a_d|R^d/2 \geq 1/\epsilon$. Therefore the restriction of $1/(f(z))$ to the circle $\{z \in \mathbb{C} \mid |z| = R\}$ has maximum modulus $\leq \epsilon$. By the Maximum Modulus Theorem,

$1/|f(z)| \leq \epsilon$ on the interior. Since this holds for every $\epsilon > 0$, it follows that $1/|f(z)|$ is identically zero on all of \mathbb{C} , which is absurd. This contradiction proves that $f(z)$ has a zero. Factoring the corresponding linear polynomial from $f(z)$ and repeating, it follows that $f(z)$ is a product of linear polynomials.

5. THE REGULAR MATRICES ARE DENSE

Definition 5.1. Let $n \geq 1$ be an integer. An $n \times n$ matrix $A \in \text{Mat}_{n \times n}(\mathbb{F})$ is *regular* if $c_A(X)$ factors as $(X - \lambda_1) \cdots (X - \lambda_n)$ for distinct elements $\lambda_1, \dots, \lambda_n \in \mathbb{F}$.

Let $(a_{i,j} | 1 \leq i, j \leq n)$ be n^2 variables. Let $f \in \mathbb{F}[a_{i,j} | 1 \leq i, j \leq n]$ be a polynomial in these variables. This defines a function $f : \text{Mat}_{n \times n}(\mathbb{F}) \rightarrow \mathbb{F}$.

Proposition 5.2. *If \mathbb{F} is infinite and algebraically closed, and if f is zero on all regular matrices, then f is the zero polynomial.*

Proof. Let A be any $n \times n$ matrix. Because \mathbb{F} is infinite, there exist distinct elements $\lambda_1, \dots, \lambda_n \in \mathbb{F}$. Let B be the matrix,

$$B(i, j) = \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j \end{cases},$$

i.e., B is the diagonal matrix with entries $\lambda_1, \dots, \lambda_n$. Let t be a variable and consider the matrix A_t with entries in $\mathbb{F}[t]$ given by,

$$A_t = tB + (1 - t)A.$$

Denote by $c_{A_t}(X)$ the polynomial $\det(XI_n - A_t)$, which is a polynomial in $\mathbb{F}[t, x]$. This polynomial is,

$$c_{A_t}(X) = X^n - \text{trace}(A_t)X^{n-1} + \cdots + (-1)^{\det(A_t)}.$$

In particular, for every value of t , this is a nonzero polynomial of degree n . By Theorem 2.6, $c_{A_t}(X)$ has a repeated linear factor iff $\text{Disc}(c_{A_t}(X))$ is zero. Now the entries of $c_{A_t}(X)$ are linear polynomials in t , thus $\text{Disc}(c_{A_t}(X))$ is a polynomial of degree at most n in t . Moreover, for $t = 1$, $c_{A_t}(X) = c_B(X) = (X - \lambda_1) \cdots (X - \lambda_n)$ has distinct linear factors. So, by Theorem 2.6, $\text{Disc}(c_{A_t}(X))$ is nonzero when $t = 1$. Hence $\text{Disc}(c_{A_t}(X))$ is not the zero polynomial in t . Therefore the polynomial $\text{Disc}(c_{A_t}(X))$ has at most n distinct zeros. So there are at most n distinct values of t for which A_t is not regular, i.e., A_t is regular for infinitely many values of t .

Consider the polynomial in t , $f(A_t)$. Unless $f(A_t)$ is the zero polynomial in t , there are only finitely many values of t where $f(A_t) = 0$. By the last paragraph, there are infinitely many values of t where A_t is regular. By hypothesis, $f(A_t) = 0$ whenever A_t is regular. So $f(A_t) = 0$ for infinitely many t , proving $f(A_t)$ is the zero polynomial in t .

In particular, plugging in $t = 0$, $A_0 = A$ and $f(A)$ is the value of $f(A_t)$ at $t = 0$. Since $f(A_t)$ is the zero polynomial, $f(A) = 0$. Since this holds for every $n \times n$ matrix A , by Proposition 3.1, f is the zero polynomial. \square

6. THE CAYLEY-HAMILTON THEOREM

The Cayley-Hamilton theorem is the key to proving Jordan canonical form. For any particular matrix, the Cayley-Hamilton theorem can be verified directly. However, the *abstract proof* that works for all matrices at once is rather involved, using all the ideas we have developed to this point. The idea is to interpret the Cayley-Hamilton theorem as a sequence of identities of polynomials in the variables $(a_{i,j})$ with integer coefficients. By Corollary 3.2, it then suffices to prove these identities are true as identities of functions on $\text{Mat}_{n \times n}(\mathbb{C})$. By the fundamental theorem of algebra and Proposition 17, it then suffices to prove these identities of functions on the set of regular matrices. This follows immediately from what we know about diagonal matrices.

Lemma 6.1. *Let $T : V \rightarrow V$ be a linear operator on a finite-dimensional \mathbb{F} -vector space. If T is diagonalizable, then $c_T(T)$ is the zero operator.*

Proof. Let \mathcal{B} be an ordered basis for T with respect to which the matrix representation is diagonal, say

$$A = [T]_{\mathcal{B},\mathcal{B}}(i,j) = \begin{cases} \lambda_i, & i = j, \\ 0, & i \neq j \end{cases}$$

Then $c_A(X) = (X - \lambda_1) \cdots (X - \lambda_n)$. Because A is a diagonal matrix, it follows easily by induction on $\deg(f)$ that for any polynomial $f(x) \in \mathbb{F}[x]$, $f(A)$ is the diagonal matrix,

$$f(A)(i,j) = \begin{cases} f(\lambda_i), & i = j, \\ 0, & i \neq j \end{cases}$$

For every $i = 1, \dots, n$, $c_A(\lambda_i) = 0$, because $X - \lambda_i$ is a factor of $c_A(X)$. Therefore $c_A(A) = 0$. Of course $[c_T(T)]_{\mathcal{B},\mathcal{B}} = c_A(A)$, thus $c_T(T)$ is the zero operator. \square

Corollary 6.2. *Let $T : V \rightarrow V$ be a linear operator that is regular, i.e., every matrix representation is regular. Then $c_T(T)$ is the zero operator.*

Proof. As proved in lecture, every regular linear operator is diagonalizable. \square

Definition 6.3. Let $n \geq 1$ be an integer and let $(a_{i,j} | 1 \leq i, j \leq n)$ be n^2 variables. Let A be the $n \times n$ matrix with entries in $\mathbb{Z}[a_{i,j} | 1 \leq i, j \leq n]$, $A(i,j) = a_{i,j}$. The *generic characteristic polynomial* is the polynomial $c_A(X) := \det(XI_n - A)$, which is a polynomial with integer coefficients in the variables $(a_{i,j} | 1 \leq i, j \leq n)$ and X . The *generic Cayley-Hamilton matrix* is the matrix with entries in $\mathbb{Z}[a_{i,j} | 1 \leq i, j \leq n]$, $c_A(A)$.

Theorem 6.4 (The Cayley-Hamilton theorem). *Every entry of the generic Cayley-Hamilton matrix is 0.*

Proof. Every entry is a polynomial in the variables $(a_{i,j} | 1 \leq i, j \leq n)$. By Corollary 3.2, it suffices to prove every entry gives the zero function on $\text{Mat}_{n \times n}(\mathbb{C})$. By Theorem 4.2, the fundamental theorem of algebra, \mathbb{C} satisfies the hypotheses of Proposition 5.2. So by Proposition 5.2, it suffices to prove that for every regular matrix $B \in \text{Mat}_{n \times n}(\mathbb{C})$, the value of the entry on B is 0. Substituting in the entries of B for the $a_{i,j}$ s, the value of $c_A(A)$ is $c_B(B)$. By Corollary 6.2, $c_B(B)$ is the zero matrix, i.e., every entry is zero. Therefore every entry of the generic Cayley-Hamilton matrix is 0. \square

Corollary 6.5. *For every field \mathbb{F} , not necessarily algebraically closed or infinite, for every integer $n \geq 1$, for every matrix $B \in \text{Mat}_{n \times n}(\mathbb{F})$, $c_B(B)$ is the zero matrix. Therefore, for every linear operator $T : V \rightarrow V$ on a finite-dimensional \mathbb{F} -vector space, $c_T(T) = 0$.*

Proof. The operation of multiplications, additions and subtractions computing $c_B(B)$ is exactly the same as the operation of multiplications, additions and subtractions computing the generic Cayley-Hamilton matrix, but with the entries of B substituted for the variables $(a_{i,j})$. Therefore $c_B(B)$ is obtained from $c_A(A)$ by substituting in the entries of B for the variables $(a_{i,j})$. By Theorem 6.4, every entry of $c_A(A)$ is the zero polynomial. Therefore the value on B is 0, i.e., $c_B(B)$ is the zero matrix. \square