

## 10.6 The Rational Canonical Form

Everything so far has needed the characteristic polynomial to split. Section 5 removed that need from the building block: a companion matrix  $\mathbf{C}(p)$  exists for every monic  $p$ , and it is the matrix of an operator on a cyclic subspace. What is left is to cut an arbitrary space into cyclic pieces. This section does that, over an **arbitrary field**, and the result is a genuine canonical form: a list of polynomials attached to  $T$ , from which the matrix is written down mechanically, and which two operators share exactly when they are similar. Along the way we learn that similarity does not change if the field is enlarged.

### 10.6.1 The plan

The goal, stated before its parts are named:

**Every operator on a nonzero finite-dimensional space decomposes the space into cyclic subspaces  $Z(v_1; T), \dots, Z(v_r; T)$  whose annihilators form a divisibility chain  $d_1 \mid d_2 \mid \dots \mid d_r = m_T$ , and that chain is determined by  $T$ .**

In the adapted basis the matrix is then  $\mathbf{C}(d_1) \oplus \dots \oplus \mathbf{C}(d_r)$ , by Theorem 10.5.2 (b) and Theorem 9.1.5. So the whole content is the decomposition, and it is the hardest theorem of the chapter. The route has three steps, and each one is a tool we already have.

- **Step 1 (one prime at a time).** Chapter 9's primary decomposition splits  $V = V_1 \oplus \dots \oplus V_k$ , where  $m_{T|_{V_i}} = p_i^{e_i}$  is a power of a single monic irreducible (Theorem 9.8.1). Nothing more is needed from the general theory.
- **Step 2 (inside one prime power).** On a piece with  $m_T = p^e$ , the operator  $P = p(T)$  is nilpotent, and the structure theorem for nilpotent operators was proved by induction, with  $\text{im } N$  as the smaller space to which the hypothesis applies (Theorem 10.2.3). We run the same induction with  $P$  in place of  $N$  and cyclic subspaces  $Z(u; T)$  in place of the spans of chains. The outcome is a decomposition into cyclic pieces whose annihilators are powers  $p^{e_1} \geq p^{e_2} \geq \dots$  of the one prime.
- **Step 3 (regroup across primes).** A cyclic piece for  $p_1$  and a cyclic piece for  $p_2$  can be **merged**: the sum of their generators has annihilator the product, by Lemma 10.5.4. Merging the largest piece of every prime gives  $d_r = m_T$ , the next largest give  $d_{r-1}$ , and so on; the divisibility chain is automatic because within one prime the exponents decrease.

Uniqueness is a count, exactly as for the nilpotent case: the dimensions  $\dim \ker p(T)^k$  are visible in the decomposition and see nothing but the exponents.

### 10.6.2 One prime at a time

Here is Step 2. Compare it with Theorem 10.2.3, which is the special case  $p = x$ : there the pieces are spanned by Jordan chains  $u, Nu, \dots$ , and here by  $u, Tu, T^2u, \dots$

#### Theorem 10.6.1. Cyclic Decomposition for a Single Prime Power

Let  $V \neq \{0\}$  be a finite-dimensional vector space over  $F$ , let  $T \in \mathcal{L}(V)$ , and suppose

$$m_T = p^e, \quad p \in F[x] \text{ monic irreducible of degree } d, \quad e \geq 1.$$

Then there are an integer  $m \geq 1$ , vectors  $u_1, \dots, u_m \in V$ , and exponents

$$e = k_1 \geq k_2 \geq \dots \geq k_m \geq 1$$

such that

$$V = Z(u_1; T) \oplus Z(u_2; T) \oplus \cdots \oplus Z(u_m; T), \quad m_{T, u_i} = p^{k_i} \quad (1 \leq i \leq m).$$

In particular  $\dim V = d(k_1 + \cdots + k_m)$ .

*Idea.* Induction on  $\dim V$ , with the smaller object found inside:  $W = \operatorname{im} p(T)$ . It is  $T$ -invariant, the minimal polynomial of  $T|_W$  is  $p^{e-1}$ , and  $\dim W < \dim V$  because  $p(T)$  is not injective. So  $W$  already breaks into cyclic pieces  $Z(w_i; T)$ . Each is one step short of what we want:  $w_i = p(T)u_i$  for some  $u_i \in V$ , and  $Z(u_i; T)$  is a longer piece, with annihilator  $p$  times that of  $w_i$ . What is still missing is counted by  $U = \ker p(T)$ , on which  $T$  has **irreducible** minimal polynomial  $p$ , so that  $U$  is a direct sum of cyclic pieces of dimension  $d$  each and no proper cyclic piece inside one of them exists — that is the case  $e = 1$ , settled first. The ends  $p^{k_i-1}(T)u_i$  of the lengthened pieces sit inside  $U$ ; complete them to a decomposition of  $U$ , and the extra generators supply the pieces with annihilator  $p$ . Then Rank–Nullity makes the dimensions add up exactly, and directness is proved by applying  $p(T)$  once, as in the nilpotent case.

*Proof.* We argue by induction on  $n = \dim V \geq 1$ , the statement being: for every field  $F$ , every monic irreducible  $p \in F[x]$ , every  $e \geq 1$ , and every  $T$  on a space of dimension  $n$  with  $m_T = p^e$ , a decomposition as described exists.

Write  $P = p(T)$  and  $U = \ker P$ , and recall that  $Z(v; T|_{V'}) = Z(v; T)$  whenever  $V'$  is a  $T$ -invariant subspace containing  $v$ , since the two operators agree on  $v$  and on everything  $T$  produces from it.

*The case  $e = 1$ .* Then  $m_T = p$  is irreducible. For  $u \neq 0$  we have  $m_{T, u} \mid m_T = p$  (Theorem 10.5.2 (d)) and  $m_{T, u} \neq 1$ , so  $m_{T, u} = p$  and  $\dim Z(u; T) = d$  by Theorem 10.5.2 (a). Two consequences.

*Every cyclic piece is minimal.* If  $W \subseteq Z(u; T)$  is  $T$ -invariant and  $W \neq \{0\}$ , pick  $w \in W$  with  $w \neq 0$ . Then  $Z(w; T) \subseteq W$  (Proposition 10.5.1 (c)) and  $\dim Z(w; T) = d = \dim Z(u; T)$ , so  $Z(w; T) = Z(u; T)$  by Theorem 1.6.8, and  $W = Z(u; T)$ . So the only  $T$ -invariant subspaces of  $Z(u; T)$  are  $\{0\}$  and  $Z(u; T)$ .

*The pieces fill  $V$ .* Among all lists  $u_1, \dots, u_m$  of nonzero vectors for which  $Z(u_1; T) + \cdots + Z(u_m; T)$  is direct, choose one with  $m$  as large as possible; this is possible because such a sum has dimension  $md \leq \dim V$  (Theorem 1.7.5). Write  $S$  for that sum. If  $S \neq V$ , pick  $v \in V \setminus S$ . The subspace  $Z(v; T) \cap S$  is  $T$ -invariant (Proposition 9.1.4), lies inside  $Z(v; T)$  and does not contain  $v$ , so it is  $\{0\}$  by the previous paragraph. Then  $S + Z(v; T)$  is direct — a relation  $c_1 + \cdots + c_m + c = 0$  with  $c_j \in Z(u_j; T)$  and  $c \in Z(v; T)$  puts  $c \in Z(v; T) \cap S = \{0\}$ , and directness of  $S$  kills the rest (Theorem 1.7.5) — contradicting the maximality of  $m$ . Hence  $V = Z(u_1; T) \oplus \cdots \oplus Z(u_m; T)$  with every  $m_{T, u_i} = p$ , and every  $k_i = 1 = e$ .

This settles  $e = 1$ , and in particular the base case  $n = 1$ , where  $m_T$  is linear.

*The case  $e \geq 2$ .* Let  $n \geq 2$  and assume the statement for all dimensions less than  $n$ .

**Step 1. Cyclic pieces inside  $\operatorname{im} P$ .** Put  $W = \operatorname{im} P$ . Since  $e \geq 2$ , the polynomial  $p$  does not annihilate  $T$  (else  $m_T = p^e$  would divide  $p$ , Theorem 9.5.1), so  $P \neq 0$  and  $W \neq \{0\}$ . Also  $P$  is not injective: otherwise  $P^e = m_T(T) = 0$  would be injective on  $V \neq \{0\}$  (Theorem 0.4.3),

which is false. So  $1 \leq \dim W < n$  by Theorem 2.3.1. The subspace  $W$  is  $T$ -invariant (Theorem 6.7.4 (b)), and

$$m_{T|_W} = p^{e-1}.$$

Indeed  $p^{e-1}(T)$  kills  $W$ , because  $p^{e-1}(T)Pv = m_T(T)v = 0$  for every  $v$ ; and  $p^{e-2}(T)$  does not, because  $p^{e-2}(T)P \neq 0$  — that operator is  $p^{e-1}(T)$ , which is nonzero since  $m_T = p^e \nmid p^{e-1}$  — so some  $Pv \in W$  survives it. Hence  $m_{T|_W}$  is a monic divisor of  $p^{e-1}$  but not of  $p^{e-2}$ , and Lemma 9.5.2 leaves only  $p^{e-1}$ . By the induction hypothesis applied to  $T|_W$ ,

$$W = Z(w_1; T) \oplus \cdots \oplus Z(w_q; T), \quad m_{T, w_i} = p^{l_i}, \quad e-1 = l_1 \geq \cdots \geq l_q \geq 1.$$

**Step 2. Lengthen each piece by one step.** Each  $w_i$  lies in  $W = \operatorname{im} P$ , so we may choose  $u_i \in V$  with  $Pu_i = w_i$ , for  $i = 1, \dots, q$ . Put  $k_i = l_i + 1$ . Then  $p^{k_i}(T)u_i = p^{l_i}(T)w_i = 0$ , so  $m_{T, u_i} \mid p^{k_i}$ ; and  $p^{l_i}(T)u_i \neq 0$ , since otherwise  $p^{l_i-1}(T)w_i = p^{l_i}(T)u_i = 0$ , contradicting  $m_{T, w_i} = p^{l_i}$ . By Lemma 9.5.2,

$$m_{T, u_i} = p^{k_i} = p^{l_i+1} \quad (1 \leq i \leq q).$$

**Step 3. Fill up  $\ker P$ .** Put  $z_i = p^{l_i}(T)u_i$  for  $i \leq q$ . Each is nonzero by Step 2, lies in  $U = \ker P$  because  $Pz_i = p^{k_i}(T)u_i = 0$ , and lies in  $Z(w_i; T)$ , since  $l_i \geq 1$  gives  $z_i = p^{l_i-1}(T)p(T)u_i = p^{l_i-1}(T)w_i$ . Consequently  $Z(z_i; T) \subseteq Z(w_i; T)$  (Proposition 10.5.1 (b), (c)), so the sum  $Z(z_1; T) + \cdots + Z(z_q; T)$  is direct: a vanishing sum of vectors  $c_i \in Z(z_i; T) \subseteq Z(w_i; T)$  forces every  $c_i = 0$  by the directness of Step 1 (Theorem 1.7.5).

Now  $U \neq \{0\}$ ,  $U$  is  $T$ -invariant (Theorem 6.7.4 (b)), and  $p(T|_U) = 0$  with  $p$  irreducible, so  $m_{T|_U} = p$  as in the case  $e = 1$ . Among all lists  $(z_1, \dots, z_q, z_{q+1}, \dots, z_s)$  of nonzero vectors of  $U$  extending  $z_1, \dots, z_q$  and with  $Z(z_1; T) + \cdots + Z(z_s; T)$  direct, choose one with  $s$  as large as possible; this is possible because each summand has dimension  $d$ , by the case  $e = 1$  applied to  $T|_U$ , so  $sd \leq \dim U$ . Let  $S$  be that direct sum. If  $S \neq U$ , pick  $y \in U \setminus S$ . The subspace  $Z(y; T) \cap S$  is  $T$ -invariant (Proposition 9.1.4), lies inside  $Z(y; T)$ , and does not contain  $y$ ; by the case  $e = 1$ , applied to  $T|_U$ , it is therefore  $\{0\}$ . Then  $S + Z(y; T)$  is direct — a relation  $c_1 + \cdots + c_s + c = 0$  with  $c_j \in Z(z_j; T)$  and  $c \in Z(y; T)$  puts  $c \in Z(y; T) \cap S = \{0\}$ , and directness of  $S$  kills the rest (Theorem 1.7.5) — contradicting the maximality of  $s$ . Hence

$$U = Z(z_1; T) \oplus \cdots \oplus Z(z_s; T), \quad m_{T, z_j} = p \text{ for every } j.$$

Put  $m = s$ , and for  $q < i \leq m$  put  $u_i = z_i$  and  $k_i = 1$ , so that  $m_{T, u_i} = p^{k_i}$  for those  $i$  as well.

**Step 4. The pieces fill  $V$ .** First count. By Theorem 10.5.2 (a),  $\dim Z(u_i; T) = \deg p^{k_i} = dk_i$ . Hence

$$\begin{aligned} \sum_{i=1}^m \dim Z(u_i; T) &= \sum_{i=1}^q d(l_i + 1) + (m - q)d \\ &= \left( \sum_{i=1}^q dl_i \right) + md = \dim W + \dim U = n, \end{aligned}$$

where  $\sum_i dl_i = \sum_i \dim Z(w_i; T) = \dim W$  by Step 1,  $md = \sum_j \dim Z(z_j; T) = \dim U$  by Step 3, and the last equality is Theorem 2.3.1.

Next, directness. Let  $y_1 + \cdots + y_m = 0$  with  $y_i \in Z(u_i; T)$ , say  $y_i = f_i(T)u_i$  (Proposition 10.5.1 (a)). Apply  $P$ . For  $i > q$  we get  $Py_i = f_i(T)Pu_i = 0$ , since  $u_i \in U$ . For  $i \leq q$  we get

$Py_i = f_i(T)w_i \in Z(w_i; T)$ . So

$$\sum_{i=1}^q f_i(T)w_i = 0,$$

and the directness of Step 1 forces  $f_i(T)w_i = 0$  for every  $i \leq q$ , that is,  $p^{l_i} \mid f_i$  (Definition 10.5.2). Write  $f_i = p^{l_i}g_i$  for  $i \leq q$ ; then

$$y_i = g_i(T)p^{l_i}(T)u_i = g_i(T)z_i \in Z(z_i; T) \quad (i \leq q).$$

For  $i > q$ ,  $y_i \in Z(u_i; T) = Z(z_i; T)$  by definition. So the relation  $y_1 + \cdots + y_m = 0$  is a relation between vectors of the independent subspaces  $Z(z_1; T), \dots, Z(z_m; T)$  of Step 3, and every  $y_i = 0$ . By Theorem 1.7.5 ((b)  $\Rightarrow$  (a)) the sum  $\sum_i Z(u_i; T)$  is direct, and by the count its dimension is  $n$ , so it is all of  $V$  (Theorem 1.6.8).

Finally the exponents. We have  $k_1 = l_1 + 1 = e$ , the  $k_i$  with  $i \leq q$  decrease because the  $l_i$  do, and the remaining ones equal  $1 \leq k_q$ ; so after listing the pieces in this order,  $e = k_1 \geq \cdots \geq k_m \geq 1$ . The dimension formula  $\dim V = d \sum_i k_i$  is the count above. This completes the induction. ■

Two sanity checks. For  $p = x$ , so that  $T$  is nilpotent with  $m_T = x^e$ , the theorem returns Theorem 10.2.3:  $Z(u; T) = \text{span}(u, Tu, \dots, T^{k-1}u)$  is the span of a Jordan chain written backwards, and  $d = 1$ , so  $\dim V = \sum_i k_i$ . And for  $e = 1$  it returns Exercise (C2: Irreducible minimal polynomials): all pieces have dimension  $d$ , so  $d \mid \dim V$ .

### 10.6.3 Invariant factors

Now Steps 1 and 3 of the plan. Merging pieces across different primes is where Lemma 10.5.4 earns its keep.

#### Theorem 10.6.2. Cyclic Decomposition Theorem

Let  $V \neq \{0\}$  be a finite-dimensional vector space over  $F$  and let  $T \in \mathcal{L}(V)$ . Then there are an integer  $r \geq 1$ , vectors  $v_1, \dots, v_r \in V$ , and monic **nonconstant** polynomials  $d_1, \dots, d_r \in F[x]$  such that

$$V = Z(v_1; T) \oplus \cdots \oplus Z(v_r; T), \quad m_{T, v_j} = d_j,$$

and

- (a)  $d_1 \mid d_2 \mid \cdots \mid d_r$ ;
- (b)  $d_r = m_T$ ;
- (c)  $d_1 d_2 \cdots d_r = p_T$ ; in particular  $\sum_j \deg d_j = \dim V$ ;
- (d) there is a **finer** decomposition  $V = \bigoplus_{i,j} Z(v_{ij}; T)$  into cyclic subspaces whose annihilators  $m_{T, v_{ij}}$  are powers of monic irreducibles, and the list of these prime powers, with repetitions, consists exactly of the prime powers occurring in the factorizations of  $d_1, \dots, d_r$ .

*Idea.* Run the plan. Primary decomposition gives one piece per prime; Theorem 10.6.1 breaks each piece into cyclic subspaces with annihilators  $p_i^{e_{i1}}, p_i^{e_{i2}}, \dots$ , the exponents decreasing. Arrange these in an array, one row per prime, with the largest exponent of each prime in the

**first** column. Multiplying down a column gives one member of the chain, and the columns are numbered **backwards**: the first column, which carries the largest exponent of every prime, gives  $d_r = m_T$ ; the second gives  $d_{r-1}$ ; the last gives  $d_1$ . The chain  $d_1 \mid \cdots \mid d_r$  comes out for free, because the exponents in a row decrease from left to right. The generator for a column is the sum of the generators in it, whose annihilator is the product by Lemma 10.5.4.

*Proof. Step 1.* By Theorem 6.3.5,  $m_T = p_1^{e_1} \cdots p_k^{e_k}$  with  $p_1, \dots, p_k$  distinct monic irreducibles and  $e_i \geq 1$ ; here  $\deg m_T \geq 1$  since  $V \neq \{0\}$  (Theorem 9.5.1). By Theorem 9.8.1,  $V = V_1 \oplus \cdots \oplus V_k$  with  $V_i$  nonzero and  $T$ -invariant and  $m_{T|_{V_i}} = p_i^{e_i}$ .

**Step 2.** Fix  $i$  and apply Theorem 10.6.1 to  $T|_{V_i}$ . Since  $Z(v; T|_{V_i}) = Z(v; T)$  for  $v \in V_i$ , it gives vectors  $v_{i1}, \dots, v_{im_i} \in V_i$  and exponents  $e_i = e_{i1} \geq \cdots \geq e_{im_i} \geq 1$  with

$$V_i = Z(v_{i1}; T) \oplus \cdots \oplus Z(v_{im_i}; T), \quad m_{T, v_{ij}} = p_i^{e_{ij}}.$$

Combining with Step 1,  $V$  is the direct sum of **all** the subspaces  $Z(v_{ij}; T)$ : a basis of  $V$  is obtained by concatenating bases of the  $V_i$ , each of which is a concatenation of bases of its own pieces, so Theorem 1.7.5 ((d)  $\Rightarrow$  (a)) applies.

**Step 3.** Let  $r = \max_i m_i$  and, for  $1 \leq j \leq r$ , put  $j' = r + 1 - j$  and

$$d_j = \prod_{i: m_i \geq j'} p_i^{e_{ij'}}, \quad v_j = \sum_{i: m_i \geq j'} v_{ij'}.$$

So  $d_r$  collects the largest exponent of every prime and  $d_1$  the smallest ones still present.

*The divisibility chain.* The exponent of  $p_i$  in  $d_j$  is  $e_{ij'}$  if  $m_i \geq j'$  and 0 otherwise. As  $j$  increases,  $j'$  decreases, so this exponent is nondecreasing in  $j$  (the exponents  $e_{i1} \geq e_{i2} \geq \cdots$  decrease in the second index, and a 0 is followed only by values  $\geq 0$ ). Hence  $d_j \mid d_{j+1}$  for every  $j$  (Lemma 9.5.2), which is (a). Taking  $j = r$ , so  $j' = 1$ , gives  $d_r = \prod_i p_i^{e_{i1}} = \prod_i p_i^{e_i} = m_T$ , which is (b). Each  $d_j$  is nonconstant: choosing  $i_0$  with  $m_{i_0} = r$ , we have  $m_{i_0} \geq j'$  for every  $j$ , so the factor  $p_{i_0}^{e_{i_0 j'}}$  occurs in  $d_j$ , with exponent at least 1.

*The annihilators and the pieces.* For fixed  $j$ , the polynomials  $p_i^{e_{ij'}}$  appearing in  $d_j$  are pairwise coprime, being powers of distinct monic irreducibles: a monic common divisor of  $p_i^a$  and  $p_{i'}^b$  with  $i \neq i'$  is a power of each, hence 1 by the uniqueness of factorization (Lemma 9.5.2, Theorem 6.3.5 (b)). So Lemma 10.5.4 gives  $m_{T, v_j} = d_j$ . Moreover

$$Z(v_j; T) = \bigoplus_{i: m_i \geq j'} Z(v_{ij'}; T) :$$

the left side is contained in the right, because the right side is a  $T$ -invariant subspace containing  $v_j$  (Proposition 9.1.4, Proposition 10.5.1 (c)), and the two have the same dimension,

$$\dim Z(v_j; T) = \deg d_j = \sum_i \deg p_i^{e_{ij'}} = \sum_i \dim Z(v_{ij'}; T),$$

by Theorem 10.5.2 (a) and Theorem 1.7.5 ((a)  $\Rightarrow$  (e)); so they are equal (Theorem 1.6.8).

*The decomposition.* Each subspace  $Z(v_{ij_0}; T)$ , with  $1 \leq j_0 \leq m_i$ , occurs in exactly one of the sums just displayed, namely the one with  $j = r + 1 - j_0$ ; so every pair  $(i, j_0)$  is used exactly

once and

$$\sum_{j=1}^r Z(v_j; T) = \sum_{i,j} Z(v_{ij}; T) = V,$$

and  $\sum_j \dim Z(v_j; T) = \sum_{i,j} \dim Z(v_{ij}; T) = \dim V$ . By Theorem 1.7.5 ((e)  $\Rightarrow$  (a)) the sum over  $j$  is direct, and it equals  $V$ .

- (c) Each  $Z(v_j; T)$  is  $T$ -invariant and nonzero, so choose a basis  $\mathcal{B}_j$  of  $Z(v_j; T)$  for each  $j$  and let  $\mathcal{B}$  be  $\mathcal{B}_1, \dots, \mathcal{B}_r$  written one after another. By Theorem 9.1.5,  $\mathcal{B}$  is a basis of  $V$  and  $[T]_{\mathcal{B}}^{\mathcal{B}} = [T|_{Z(v_1; T)}]_{\mathcal{B}_1}^{\mathcal{B}_1} \oplus \dots \oplus [T|_{Z(v_r; T)}]_{\mathcal{B}_r}^{\mathcal{B}_r}$ . Then  $x\mathbf{I} - [T]_{\mathcal{B}}^{\mathcal{B}}$  is block diagonal with the blocks  $x\mathbf{I} - [T|_{Z(v_j; T)}]_{\mathcal{B}_j}^{\mathcal{B}_j}$ , so Theorem 7.8.4, applied  $r - 1$  times, gives

$$p_T = \det (x\mathbf{I} - [T]_{\mathcal{B}}^{\mathcal{B}}) = \prod_{j=1}^r p_{T|_{Z(v_j; T)}}$$

(Definition 7.10.2). By Theorem 10.5.2 (c),  $p_{T|_{Z(v_j; T)}} = m_{T, v_j} = d_j$ . Hence  $p_T = d_1 \cdots d_r$ , and comparing degrees gives  $\sum_j \deg d_j = \deg p_T = \dim V$  (Theorem 7.10.2).

- (d) Step 2 already produced the finer decomposition  $V = \bigoplus_{i,j} Z(v_{ij}; T)$ , with  $m_{T, v_{ij}} = p_i^{e_{ij}}$  a prime power. By the definition of  $d_j$  in Step 3, the exponent of  $p_i$  in  $d_j$  is  $e_{i, r+1-j}$  when  $m_i \geq r + 1 - j$  and 0 otherwise, and each pair  $(i, j)$  with  $1 \leq j \leq m_i$  is used exactly once. So the prime powers occurring in  $d_1, \dots, d_r$  are exactly the polynomials  $p_i^{e_{ij}}$ , with the same repetitions. ■

The polynomials in the chain, and the prime powers they are built from, are the data we are after — but first we must know that they do not depend on the decomposition we happened to construct. The count that settles this is the count of Section 2 in disguise: there the chain lengths were read off from the ranks of the powers of  $N$ , and here the exponents are read off from the dimensions of the kernels of the powers of  $p(T)$ . Write  $v_p(f)$  for the exponent of a monic irreducible  $p$  in the factorization of a nonzero  $f \in F[x]$ , so that  $v_p(f) \geq 1$  exactly when  $p \mid f$ .

### Proposition 10.6.3. Counting the Exponents

Let  $V \neq \{0\}$  be finite-dimensional over  $F$ , let  $T \in \mathcal{L}(V)$ , and suppose

$$V = Z(u_1; T) \oplus \dots \oplus Z(u_N; T), \quad m_{T, u_i} = g_i,$$

with every  $g_i$  monic and nonconstant. Then for every monic irreducible  $p \in F[x]$ :

(a)  $\dim \ker p(T)^k = \deg p \cdot \sum_{i=1}^N \min(k, v_p(g_i))$  for every integer  $k \geq 0$ ;

(b) for every  $k \geq 1$ ,

$$\#\{i : v_p(g_i) \geq k\} = \frac{\dim \ker p(T)^k - \dim \ker p(T)^{k-1}}{\deg p},$$

a number depending only on  $T$ ,  $p$ , and  $k$ .

*Idea.* Kernels respect the decomposition, so we may work inside one piece. There a vector is  $h(T)u$ , and it is killed by  $p(T)^k$  exactly when  $g \mid p^k h$ , which after canceling the common factor  $\gcd(g, p^k)$  says that  $h$  is a multiple of  $g/\gcd(g, p^k)$ . The surviving vectors are therefore a cyclic subspace again, generated by that multiple of  $u$ , and its annihilator is  $\gcd(g, p^k) = p^{\min(k, v_p(g))}$ . Summing over the pieces and taking differences in  $k$  then counts exponents.

*Proof.*

- (a) Write  $Z_i = Z(u_i; T)$  and  $S = p(T)^k$ . Each  $Z_i$  is  $T$ -invariant, hence  $S$ -invariant (Theorem 6.7.4 (b)). If  $v = v_1 + \cdots + v_N$  with  $v_i \in Z_i$ , then  $Sv = Sv_1 + \cdots + Sv_N$  with  $Sv_i \in Z_i$ , so  $Sv = 0$  if and only if every  $Sv_i = 0$ , by the uniqueness of the splitting (Theorem 1.7.5). Therefore

$$\ker S = (\ker S \cap Z_1) \oplus \cdots \oplus (\ker S \cap Z_N),$$

and it suffices to compute each  $\dim(\ker S \cap Z_i)$ . Fix  $i$  and write  $u = u_i$ ,  $g = g_i$ ,  $Z = Z_i$ . Put

$$c = \gcd(g, p^k) = p^{\min(k, v_p(g))} \quad \text{and} \quad t = g/c,$$

the formula for  $c$  holding because a monic common divisor of  $g$  and  $p^k$  is a power of  $p$  (Lemma 9.5.2) dividing both, and  $p^s \mid g$  exactly when  $s \leq v_p(g)$  (Theorem 6.3.5). Every element of  $Z$  is  $h(T)u$  for some  $h \in F[x]$  (Proposition 10.5.1 (a)), and

$$p(T)^k h(T)u = 0 \iff g \mid p^k h \iff t \mid h,$$

where the second equivalence holds because  $g = ct$  and  $p^k = c(p^k/c)$  with  $\gcd(t, p^k/c) = 1$ : from  $ct \mid c(p^k/c)h$  we get  $t \mid (p^k/c)h$ , hence  $t \mid h$  (Theorem 6.2.8), and conversely  $t \mid h$  gives  $g = ct \mid ch \mid p^k h$  since  $c \mid p^k$ . So

$$\ker S \cap Z = \{h(T)u : t \mid h\} = \{s(T)(t(T)u) : s \in F[x]\} = Z(t(T)u; T).$$

Its generator has annihilator  $c$ : for  $s \in F[x]$ ,  $s(T)t(T)u = 0$  if and only if  $g \mid st$ , that is,  $c \mid s$ , because  $g = ct$ . Taking  $s = 1$  shows that  $t(T)u \neq 0$  unless  $c = 1$ , so Theorem 10.5.2 (a) applies whenever  $c \neq 1$  and gives

$$\dim(\ker S \cap Z_i) = \deg c = \min(k, v_p(g_i)) \deg p,$$

where for  $c = 1$  the cyclic subspace is  $\{0\}$  of dimension 0, in agreement. Adding over  $i$  gives (a).

- (b) Subtract the values of (a) at  $k$  and  $k-1$ . Each term  $\min(k, v_p(g_i)) - \min(k-1, v_p(g_i))$  equals 1 if  $v_p(g_i) \geq k$  and 0 otherwise, so the difference is  $\deg p$  times the number of indices  $i$  with  $v_p(g_i) \geq k$ . The left-hand side is defined from  $T$ ,  $p$ , and  $k$  alone. ■

#### Corollary 10.6.4. Uniqueness of the Chain and of the Prime Powers

Let  $V \neq \{0\}$  be finite-dimensional over  $F$  and  $T \in \mathcal{L}(V)$ .

- (a) Any two decompositions as in Theorem 10.6.2 give the **same** ordered list  $d_1 \mid \cdots \mid d_r$ .

- (b) Any decomposition  $V = Z(u_1; T) \oplus \cdots \oplus Z(u_N; T)$  whose annihilators  $m_{T, u_i}$  are powers of monic irreducibles gives the same list of annihilators, up to order, namely the prime powers occurring in  $d_1, \dots, d_r$ .
- (c) Explicitly, writing

$$c_p(k) = \frac{\dim \ker p(T)^k - \dim \ker p(T)^{k-1}}{\deg p} \quad (p \text{ monic irreducible, } k \geq 1),$$

the list is recovered by  $r = \max_p c_p(1)$  and

$$v_p(d_j) = \#\{k \geq 1 : c_p(k) \geq r + 1 - j\} \quad (1 \leq j \leq r).$$

In particular the invariant factors of  $T$  are determined by the numbers  $\dim \ker p(T)^k$  alone.

*Idea.* Divisibility puts the zeros first: in the list  $v_p(d_1) \leq \cdots \leq v_p(d_r)$ , the indices where  $p$  actually occurs are a block at the **right** end. So a count of how many  $d_j$  are divisible by  $p^k$  pins down *which* ones they are, and therefore each individual exponent. The counts are supplied by [Proposition 10.6.3](#), which sees only  $T$ .

*Proof.*

- (a) and (c). Let  $V = \bigoplus_{j=1}^r Z(v_j; T)$  be such a decomposition, with annihilators  $d_1 \mid \cdots \mid d_r$ , all monic and nonconstant. Fix a monic irreducible  $p$ . By [Proposition 10.6.3](#) (b), applied to this decomposition,

$$c_p(k) = \#\{j : v_p(d_j) \geq k\} \quad (k \geq 1),$$

so the left-hand side, which is defined from  $T$ ,  $p$ , and  $k$  alone, computes the right-hand side. Divisibility makes the exponents nondecreasing,  $v_p(d_1) \leq \cdots \leq v_p(d_r)$ , so for each  $k$  the set  $\{j : v_p(d_j) \geq k\}$  is the set of the **last**  $c_p(k)$  indices:

$$v_p(d_j) \geq k \iff j \geq r + 1 - c_p(k) \iff c_p(k) \geq r + 1 - j. \quad (*)$$

First,  $r$  is determined. Each  $d_j$  is nonconstant, so  $v_p(d_1) \geq 1$  for some monic irreducible  $p$ ; for that  $p$ ,  $(*)$  with  $k = 1$  and  $j = 1$  gives  $c_p(1) \geq r$ , while  $c_p(1) \leq r$  always. Hence  $r = \max_p c_p(1)$ . Next, for each  $p$  and each  $j$ , counting the  $k \geq 1$  for which  $(*)$  holds gives

$$v_p(d_j) = \#\{k \geq 1 : c_p(k) \geq r + 1 - j\},$$

a finite number, since  $c_p(k) = 0$  for  $k > \deg m_T$ . This proves (c). Finally each  $d_j$  is monic with determined exponents  $v_p(d_j)$  for every  $p$ , so  $d_j$  is determined ([Theorem 6.3.5](#)), which proves (a).

- (b) Let  $V = \bigoplus_i Z(u_i; T)$  with  $m_{T, u_i} = q_i^{f_i}$ ,  $q_i$  monic irreducible, and  $f_i \geq 1$ . For a monic irreducible  $p$  we have  $v_p(q_i^{f_i}) = f_i$  if  $q_i = p$  and 0 otherwise. Applying [Proposition 10.6.3](#) (b) to this decomposition and to the one in (a) gives, for every  $p$  and every  $k \geq 1$ ,

$$\#\{i : q_i = p, f_i \geq k\} = c_p(k) = \#\{j : v_p(d_j) \geq k\}.$$

Hence, for each  $p$ , the exponents  $f_i$  with  $q_i = p$  and the positive exponents  $v_p(d_j)$  form the same multiset. So the list  $(q_i^{f_i})_i$  is exactly the list of prime powers occurring in

$d_1, \dots, d_r$ , with the same repetitions. By [Theorem 10.6.2](#) (d) such a decomposition exists, so both lists are nonempty and agree. ■

### Definition 10.6.1. Invariant Factors, Elementary Divisors

Let  $V \neq \{0\}$  be finite-dimensional over  $F$  and  $T \in \mathcal{L}(V)$ . The **invariant factors** of  $T$  are the monic nonconstant polynomials

$$d_1 \mid d_2 \mid \cdots \mid d_r$$

of [Theorem 10.6.2](#), which by [Corollary 10.6.4](#) (a) do not depend on the decomposition. The **elementary divisors** of  $T$  are the prime powers occurring in the factorizations of  $d_1, \dots, d_r$ , listed with repetition; by [Corollary 10.6.4](#) (b) they are the annihilators of the pieces in any decomposition of  $V$  into cyclic subspaces with prime-power annihilators. For  $\mathbf{A} \in M_n(F)$  the invariant factors and elementary divisors are those of  $T_{\mathbf{A}}$ .

In words: the invariant factors are the annihilators of the cyclic pieces, ordered so that each divides the next; the largest is  $m_T$ , and their product is  $p_T$ . The elementary divisors are the same information cut along the primes. Each list determines the other: one by factoring, the other by the regrouping recipe of Step 3 in the proof of [Theorem 10.6.2](#), which [Corollary 10.6.4](#) (c) shows to be forced.

**Nonexample by minimal change.** Drop the divisibility clause and keep only the product. For an operator on a  $\mathbb{Q}$ -space of dimension 3 with  $p_T = (x-1)(x-2)^2$ , the pair

$$d_1 = x-1, \quad d_2 = (x-2)^2$$

multiplies to  $p_T$  and consists of monic nonconstant polynomials of the right total degree. It is still not a list of invariant factors of anything, because  $d_1 \nmid d_2$ . The clause that fails is (a) of [Theorem 10.6.2](#), and with it (b): the largest member of the list has to be  $m_T$ , and  $m_T(T)$  kills every vector of  $V$ , so  $m_T$  is a multiple of the annihilator of every piece ([Theorem 10.5.2](#) (d)). The genuine lists here are  $((x-1)(x-2)^2)$  and  $(x-2, (x-1)(x-2))$ .

*Quick check.* Let  $T$  be an operator on a  $\mathbb{Q}$ -vector space with  $p_T = (x^2+1)^2$ . How many possibilities are there for the list of elementary divisors, and what are the corresponding invariant factors?

*Solution.* The product of the elementary divisors is  $p_T$  (by [Theorem 10.6.2](#) (c) and [Definition 10.6.1](#)), and  $x^2+1$  is irreducible over  $\mathbb{Q}$  (no rational root, [Theorem 6.3.1](#) (b)). So the elementary divisors are powers of  $x^2+1$  with exponents forming a partition of 2: either  $(x^2+1)^2$  alone, or  $x^2+1$  twice. The invariant factors are  $d_1 = (x^2+1)^2$  in the first case (one piece,  $r=1$ ) and  $d_1 = d_2 = x^2+1$  in the second. The two are told apart by  $m_T$ , which is  $d_r$ :  $(x^2+1)^2$  against  $x^2+1$ .

## 10.6.4 The rational canonical form

Everything is now assembled. Writing the matrix down is Theorem 10.5.2 (b), one piece at a time.

*Every operator over every field has a block diagonal matrix whose blocks are companion matrices of a divisibility chain, and the chain is unique.*

### Theorem 10.6.5. Rational Canonical Form

Let  $V \neq \{0\}$  be a finite-dimensional vector space over **any** field  $F$ , and let  $T \in \mathcal{L}(V)$  have invariant factors  $d_1 \mid d_2 \mid \cdots \mid d_r$ . Then there is a basis  $\mathcal{B}$  of  $V$  with

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \mathbf{C}(d_1) \oplus \mathbf{C}(d_2) \oplus \cdots \oplus \mathbf{C}(d_r),$$

and this matrix is determined by  $T$ . It is called the **rational canonical form** of  $T$ . Equivalently, every  $\mathbf{A} \in M_n(F)$  with  $n \geq 1$  is similar to exactly one matrix of the form  $\mathbf{C}(d_1) \oplus \cdots \oplus \mathbf{C}(d_r)$  with  $d_1 \mid \cdots \mid d_r$  monic and nonconstant.

*Proof.* By Theorem 10.6.2,  $V = Z(v_1; T) \oplus \cdots \oplus Z(v_r; T)$  with  $m_{T, v_j} = d_j$ . Each piece is  $T$ -invariant and nonzero, and by Theorem 10.5.2 (a), (b) the list  $\mathcal{B}_j = (v_j, Tv_j, \dots, T^{\deg d_j - 1} v_j)$  is a basis of  $Z(v_j; T)$  in which the matrix of the restriction is  $\mathbf{C}(d_j)$ . Let  $\mathcal{B}$  be  $\mathcal{B}_1, \dots, \mathcal{B}_r$  written one after another; by Theorem 9.1.5,  $\mathcal{B}$  is a basis of  $V$  and  $[T]_{\mathcal{B}}^{\mathcal{B}} = \mathbf{C}(d_1) \oplus \cdots \oplus \mathbf{C}(d_r)$ . The blocks are determined by  $T$ , since the  $d_j$  are (Corollary 10.6.4 (a)).

For a matrix  $\mathbf{A} \in M_n(F)$ , apply this to  $T_{\mathbf{A}}$  and use Theorem 4.2.4. For the uniqueness in the matrix form, suppose  $\mathbf{A} \sim \mathbf{C}(f_1) \oplus \cdots \oplus \mathbf{C}(f_s)$  with  $f_1 \mid \cdots \mid f_s$  monic and nonconstant. By Theorem 4.2.4 there is a basis  $\mathcal{C}$  of  $F^n$  with  $[T_{\mathbf{A}}]_{\mathcal{C}}^{\mathcal{C}} = \mathbf{C}(f_1) \oplus \cdots \oplus \mathbf{C}(f_s)$ . Let  $\mathcal{C}_j$  be the segment of  $\mathcal{C}$  belonging to the  $j$ -th block and let  $U_j$  be its span; then  $F^n = U_1 \oplus \cdots \oplus U_s$  with each  $U_j$  invariant and  $[T_{\mathbf{A}}|_{U_j}]_{\mathcal{C}_j}^{\mathcal{C}_j} = \mathbf{C}(f_j)$  (Theorem 9.1.5). By Theorem 10.5.3, Theorem 9.5.5, and Definition 7.10.2,  $m_{T_{\mathbf{A}}|_{U_j}} = p_{T_{\mathbf{A}}|_{U_j}} = f_j$ , so  $T_{\mathbf{A}}|_{U_j}$  is cyclic (Theorem 10.5.6) and  $U_j = Z(\mathbf{u}_j; T_{\mathbf{A}})$  for some  $\mathbf{u}_j$  with  $m_{T_{\mathbf{A}}, \mathbf{u}_j} = f_j$  (Theorem 10.5.2 (c)). This is a decomposition as in Theorem 10.6.2, so  $(f_1, \dots, f_s) = (d_1, \dots, d_r)$  by Corollary 10.6.4 (a). Hence the matrix is the rational canonical form of  $\mathbf{A}$ . ■

This pays off the promise made in Section 3: **there is a canonical form over every field**. No hypothesis on  $F$  was used, and none is needed. The form is canonical in the strong sense: the blocks come in a prescribed order, dictated by divisibility, so two operators have the *same* form, not merely similar forms.

### Corollary 10.6.6. Similarity via the Invariant Factors

Let  $\mathbf{A}, \mathbf{B} \in M_n(F)$  with  $n \geq 1$ . The following are equivalent:

- (a)  $\mathbf{A} \sim \mathbf{B}$ ;
- (b)  $\mathbf{A}$  and  $\mathbf{B}$  have the same invariant factors;

- (c)  $\mathbf{A}$  and  $\mathbf{B}$  have the same elementary divisors, up to order;  
 (d)  $\dim \ker p(\mathbf{A})^k = \dim \ker p(\mathbf{B})^k$  for every monic irreducible  $p \in F[x]$  and every  $k \geq 1$ .

*Proof.*

- (a)  $\Rightarrow$  (d). If  $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ , then  $p(\mathbf{B})^k = \mathbf{P}^{-1}p(\mathbf{A})^k\mathbf{P}$  (Proposition 4.2.6 (c)), so the two have the same rank (Proposition 4.2.6 (a)) and hence the same nullity (Theorem 3.5.7).  
 (b)  $\Rightarrow$  (b). By Corollary 10.6.4 (c), the invariant factors of a matrix are computed from the numbers  $\dim \ker p(\cdot)^k$  alone: those numbers give the counts  $c_p(k)$ , which give the length  $r$  and the exponents  $v_p(d_j)$ . Equal numbers therefore give equal invariant factors.  
 (c)  $\Leftrightarrow$  (c). The elementary divisors are the prime powers in the factorizations of the invariant factors (Definition 10.6.1), so (b)  $\Rightarrow$  (c). Conversely,  $c_p(k)$  is the number of elementary divisors of the form  $p^f$  with  $f \geq k$  (Corollary 10.6.4 (b) and (c)), so the list of elementary divisors determines every  $c_p(k)$ , hence the invariant factors by Corollary 10.6.4 (c); so (c)  $\Rightarrow$  (b).  
 (d)  $\Rightarrow$  (a). Both matrices are similar to  $\mathbf{C}(d_1) \oplus \cdots \oplus \mathbf{C}(d_r)$  (Theorem 10.6.5), and similarity is an equivalence relation (Example (Similarity)).

■

So the invariant factors are a **complete set of similarity invariants over every field**, which is the promise of Chapter 7 kept in full; Corollary 10.3.3 was the special case in which  $p_{\mathbf{A}}$  splits. Counting similarity classes becomes counting divisibility chains: over any field, the classes of  $n \times n$  matrices with a given characteristic polynomial  $p$  correspond to the ways of writing  $p = d_1 \cdots d_r$  with  $d_1 \mid \cdots \mid d_r$  monic and nonconstant.

*Warning.* The invariant factors are not the factors of  $p_{\mathbf{A}}$  in any naive sense, and  $m_{\mathbf{A}}$  alone is only the last of them. For  $\mathbf{A} = \text{diag}(1, 1, 2)$  over  $\mathbb{Q}$  the invariant factors are  $d_1 = x - 1$  and  $d_2 = (x - 1)(x - 2) = m_{\mathbf{A}}$ , and  $d_1 d_2 = (x - 1)^2(x - 2) = p_{\mathbf{A}}$ ; the factorization of  $p_{\mathbf{A}}$  into linear factors plays no role in the chain. Nor does the chain have length  $\deg p_{\mathbf{A}}$ : here  $r = 2$  for a  $3 \times 3$  matrix, and  $r = 1$  exactly when  $\mathbf{A}$  is cyclic (Theorem 10.5.6).

## 10.6.5 Elementary divisors, and the way back to Jordan

Cutting the pieces as finely as possible gives a second form, usually with more and smaller blocks.

### Theorem 10.6.7. Elementary Divisor Form

Let  $V \neq \{0\}$  be finite-dimensional over  $F$ , let  $T \in \mathcal{L}(V)$ , and let  $q_1, \dots, q_N$  be the elementary divisors of  $T$ . Then there is a basis  $\mathcal{B}$  of  $V$  with

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \mathbf{C}(q_1) \oplus \mathbf{C}(q_2) \oplus \cdots \oplus \mathbf{C}(q_N),$$

and the blocks are determined by  $T$  up to their order.

*Proof.* By [Theorem 10.6.2](#) (d) there is a decomposition  $V = \bigoplus_{i,j} Z(v_{ij}; T)$  into cyclic subspaces whose annihilators are prime powers, and by [Corollary 10.6.4](#) (b) those annihilators are exactly the elementary divisors  $q_1, \dots, q_N$  of  $T$ , up to order. Take the basis of [Theorem 10.5.2](#) (a) in each piece and concatenate; [Theorem 10.5.2](#) (b) and [Theorem 9.1.5](#) give the displayed matrix. The list of blocks depends on  $T$  alone, again by [Corollary 10.6.4](#) (b). ■

When the characteristic polynomial splits, this is the Jordan form in disguise, because a companion block for a power of a linear polynomial is a Jordan block.

### Corollary 10.6.8. Companion Blocks of Linear Powers Are Jordan Blocks

Let  $\lambda \in F$  and  $k \geq 1$ . Then

$$\mathbf{C}((x - \lambda)^k) \sim \mathbf{J}_k(\lambda).$$

Consequently, if  $p_T$  splits over  $F$ , then the elementary divisors of  $T$  are the polynomials  $(x - \lambda)^k$  for the Jordan blocks  $\mathbf{J}_k(\lambda)$  of  $T$ , and the elementary divisor form of [Theorem 10.6.7](#) and the Jordan form of  $T$  are obtained from one another by replacing each block  $\mathbf{C}((x - \lambda)^k)$  by  $\mathbf{J}_k(\lambda)$  and conversely.

*Proof.* Put  $\mathbf{J} = \mathbf{J}_k(\lambda)$ . Then  $\mathbf{J} - \lambda\mathbf{I} = \mathbf{J}_k(0)$  satisfies  $\mathbf{J}_k(0)^k = \mathbf{O} \neq \mathbf{J}_k(0)^{k-1}$ , so  $(x - \lambda)^k$  annihilates  $\mathbf{J}$  and no smaller power does; by [Lemma 9.5.2](#) and [Theorem 9.5.1](#),  $m_{\mathbf{J}} = (x - \lambda)^k$ . Also  $p_{\mathbf{J}} = (x - \lambda)^k$  ([Theorem 7.4.6](#)), so  $m_{\mathbf{J}} = p_{\mathbf{J}}$  and  $\mathbf{J}$  is cyclic ([Theorem 10.5.6](#)). By the last sentence of that theorem,  $\mathbf{J}$  is similar to  $\mathbf{C}(m_{\mathbf{J}}) = \mathbf{C}((x - \lambda)^k)$ .

Now suppose  $p_T$  splits and let  $\mathcal{B}$  be a Jordan basis, with  $[T]_{\mathcal{B}}^{\mathcal{B}} = \mathbf{J}_{k_1}(\mu_1) \oplus \dots \oplus \mathbf{J}_{k_m}(\mu_m)$  ([Theorem 10.3.1](#)). Each block corresponds to a  $T$ -invariant subspace  $U_l$ , spanned by the segment  $\mathcal{B}_l$  of  $\mathcal{B}$  belonging to it, with  $V = U_1 \oplus \dots \oplus U_m$  and  $[T|_{U_l}]_{\mathcal{B}_l}^{\mathcal{B}_l} = \mathbf{J}_{k_l}(\mu_l)$ . By the first paragraph,  $m_{T|_{U_l}} = p_{T|_{U_l}} = (x - \mu_l)^{k_l}$ , so  $T|_{U_l}$  is cyclic: there is  $v_l \in U_l$  with  $U_l = Z(v_l; T)$  and  $m_{T, v_l} = (x - \mu_l)^{k_l}$  ([Theorem 10.5.6](#), [Theorem 10.5.2](#) (c)). Since  $x - \mu_l$  is irreducible, being of degree 1, this exhibits  $V$  as a direct sum of cyclic subspaces with prime-power annihilators, so by [Corollary 10.6.4](#) (b) the elementary divisors of  $T$  are exactly the polynomials  $(x - \mu_l)^{k_l}$ . Replacing each Jordan block by the similar companion block, or conversely, therefore turns one form into the other. ■

*Remark.* The two forms carry the same information in different shapes, and each is better for something. The rational canonical form of [Theorem 10.6.5](#) has the fewest blocks and reads off  $m_T = d_r$  at a glance; the elementary divisor form has the most blocks and reads off the primes, and over a splitting field it is the Jordan form. The rational form is also called the **Frobenius normal form**. The word *rational* refers to the field: the form is computed by rational operations in  $F$ , with no root extraction, which is exactly why it exists over every field.

### 10.6.6 Similarity does not depend on the field

Here is a striking consequence, and one that would be hard to guess. Two rational matrices might conceivably be conjugate by a complex matrix and by no rational one. They cannot be.

#### Lemma 10.6.9. Rank Does Not Change with the Field

Let  $K$  be a subfield of a field  $L$  and let  $\mathbf{M} \in M_{m \times n}(K)$ . Then the rank of  $\mathbf{M}$  computed over  $K$  equals its rank computed over  $L$ .

*Proof.* If  $\mathbf{M} = \mathbf{O}$ , both ranks are 0. Otherwise, by Theorem 7.7.1 (a), each rank is the largest  $k$  such that some  $k \times k$  minor of  $\mathbf{M}$  is nonzero. The minors are determinants of submatrices of  $\mathbf{M}$ , computed by the Leibniz formula from the entries of  $\mathbf{M}$  (Definition 7.4.1), so they are the **same elements of  $K$**  whichever field we regard  $\mathbf{M}$  in. And an element of  $K$  is zero in  $K$  if and only if it is zero in  $L$ . So the two largest such  $k$  agree. ■

#### Corollary 10.6.10. Similarity Is Independent of the Field

Let  $K$  be a subfield of  $L$  and let  $\mathbf{A}, \mathbf{B} \in M_n(K)$ ,  $n \geq 1$ . If  $\mathbf{A}$  and  $\mathbf{B}$  are similar over  $L$ , then they are similar over  $K$ :

$$\begin{aligned} \mathbf{A} &= \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q} \text{ for some invertible } \mathbf{Q} \in M_n(L) \\ \implies \mathbf{A} &= \mathbf{P}^{-1}\mathbf{B}\mathbf{P} \text{ for some invertible } \mathbf{P} \in M_n(K). \end{aligned}$$

*Idea.* Similarity over  $K$  is decided by the numbers  $\dim \ker p(\mathbf{A})^k$  for the monic irreducibles  $p$  of  $K[x]$  (Corollary 10.6.6). Each such number is a nullity of a matrix with entries in  $K$ , and nullities do not notice a larger field. Similarity over  $L$  makes the numbers agree when computed over  $L$ , hence over  $K$ .

*Proof.* Let  $p \in K[x]$  be monic irreducible and  $k \geq 1$ . The matrices  $p(\mathbf{A})^k$  and  $p(\mathbf{B})^k$  have entries in  $K$ . Since  $\mathbf{A}$  and  $\mathbf{B}$  are similar over  $L$ , say  $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q}$ , we have  $p(\mathbf{A})^k = \mathbf{Q}^{-1}p(\mathbf{B})^k\mathbf{Q}$  (Proposition 4.2.6 (c)), so these two matrices have the same rank over  $L$  (Proposition 4.2.6 (a)). By Lemma 10.6.9 their ranks over  $K$  are the same numbers, hence

$$\dim \ker p(\mathbf{A})^k = n - \text{rank } p(\mathbf{A})^k = n - \text{rank } p(\mathbf{B})^k = \dim \ker p(\mathbf{B})^k$$

over  $K$ , by Theorem 3.5.7. As  $p$  and  $k$  were arbitrary, Corollary 10.6.6 ((d)  $\Rightarrow$  (a)) gives  $\mathbf{A} \sim \mathbf{B}$  over  $K$ . ■

The invariant factors are computed in  $K[x]$  and never need a root; that is the whole reason. Note what the corollary does **not** say: the invariant factors themselves do change with the field, since the irreducible polynomials do. Over  $\mathbb{Q}$  the matrix  $\mathbf{C}((x^2 + 1)^2)$  has the single invariant factor

$(x^2 + 1)^2$  and the single elementary divisor  $(x^2 + 1)^2$ ; over  $\mathbb{C}$  its invariant factor is unchanged, but its elementary divisors are  $(x - i)^2$  and  $(x + i)^2$ , and its Jordan form is  $\mathbf{J}_2(i) \oplus \mathbf{J}_2(-i)$ . What is field-independent is the **answer to the similarity question**, not the bookkeeping.

## 10.6.7 Computing the forms

*Algorithm.* **Invariant factors, elementary divisors, and the two forms of  $\mathbf{A} \in M_n(F)$ .**

1. Compute  $p_{\mathbf{A}}$  and factor it into monic irreducibles over  $F$ :  $p_{\mathbf{A}} = p_1^{a_1} \cdots p_k^{a_k}$ .
2. For each  $i$  and each  $k \geq 1$ , compute  $\dim \ker p_i(\mathbf{A})^k$  until it stops growing, and read off the exponents of the elementary divisors for  $p_i$  from [Proposition 10.6.3](#), subtracting twice for exact counts. (Equivalently: find  $m_{\mathbf{A}}$  and, in small cases, use  $\sum_{i,j} e_{ij} \deg p_i = n$  with  $\max_j e_{ij} =$  the exponent of  $p_i$  in  $m_{\mathbf{A}}$ .)
3. Group the exponents of each prime into a row, in decreasing order, and multiply down the columns, numbering them backwards: the first column gives  $d_r = m_{\mathbf{A}}$ , the second  $d_{r-1}$ , the last  $d_1$ .
4. Write  $\mathbf{C}(d_1) \oplus \cdots \oplus \mathbf{C}(d_r)$ , or  $\bigoplus_{i,j} \mathbf{C}(p_i^{e_{ij}})$ .
5. For a basis: find a cyclic generator for each piece, as in Section 5.

### Example. Invariant Factors from Polynomial Data

Let  $T$  be an operator on a 6-dimensional  $\mathbb{Q}$ -vector space with

$$p_T = (x - 1)^2(x^2 + 1)^2, \quad m_T = (x - 1)(x^2 + 1)^2.$$

Find the elementary divisors, the invariant factors, the rational canonical form ([Theorem 10.6.5](#)), and the elementary divisor form ([Theorem 10.6.7](#)).

*Solution. The primes.*  $x - 1$  and  $x^2 + 1$  are monic irreducible over  $\mathbb{Q}$  (the second has no rational root, [Theorem 6.3.1 \(b\)](#)), and  $\deg p_T = 2 + 4 = 6$  as it must be.

*The exponents.* The elementary divisors are powers  $(x - 1)^a$  and  $(x^2 + 1)^b$ . For each prime, the largest exponent occurring is its exponent in  $m_T = d_r$  ([Theorem 10.6.2 \(b\)](#), [Definition 10.6.1](#)), and the exponents weighted by the degrees sum to  $\deg p_T$  ([Theorem 10.6.2 \(c\)](#)). For  $x - 1$ : the largest exponent is 1 and the total contribution is 2, so there are two elementary divisors  $x - 1$ . For  $x^2 + 1$ : the largest exponent is 2 and the total contribution is  $4 = 2 \cdot 2$ , so there is exactly one elementary divisor  $(x^2 + 1)^2$ . Hence the elementary divisors are

$$x - 1, \quad x - 1, \quad (x^2 + 1)^2.$$

*The invariant factors.* Arrange the exponents by prime in decreasing order:

$$x - 1 : 1, 1; \quad x^2 + 1 : 2.$$

So  $r = \max(2, 1) = 2$ , and multiplying down the columns, the first giving  $d_2$  and the second  $d_1$ ,

$$d_2 = (x - 1)(x^2 + 1)^2 = m_T, \quad d_1 = x - 1.$$

Check:  $d_1 \mid d_2$  and  $d_1 d_2 = (x - 1)^2(x^2 + 1)^2 = p_T$ .

*The forms.* With  $d_2 = (x-1)(x^2+1)^2 = (x-1)(x^4+2x^2+1) = x^5 - x^4 + 2x^3 - 2x^2 + x - 1$ ,

$$\mathbf{C}(d_1) \oplus \mathbf{C}(d_2) = (1) \oplus \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

of total size  $1 + 5 = 6$ . The elementary divisor form is smaller-blocked:

$$\mathbf{C}(x-1) \oplus \mathbf{C}(x-1) \oplus \mathbf{C}((x^2+1)^2) = (1) \oplus (1) \oplus \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

of total size  $1 + 1 + 4 = 6$ . Read as a complex matrix, the second form splits further: over  $\mathbb{C}$  we have  $x^2+1 = (x-i)(x+i)$ , so the elementary divisors become  $x-1, x-1, (x-i)^2, (x+i)^2$  — the invariant factors  $d_1, d_2$  are unchanged — and [Corollary 10.6.8](#) turns the elementary divisor form into the Jordan form  $(1) \oplus (1) \oplus \mathbf{J}_2(i) \oplus \mathbf{J}_2(-i)$ . Over  $\mathbb{Q}$  no further splitting is possible, since  $x^2+1$  is irreducible there.

### Example. The Rational Canonical Form of a Matrix over the Rationals

Let

$$\mathbf{A} = \begin{pmatrix} 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in M_4(\mathbb{Q}), \quad p_{\mathbf{A}} = (x^2+1)^2.$$

Find the invariant factors of  $\mathbf{A}$ , its rational canonical form ([Theorem 10.6.5](#)), and an invertible  $\mathbf{P} \in M_4(\mathbb{Q})$  realizing it.

*Solution. The characteristic polynomial.* In  $2 \times 2$  blocks,  $\mathbf{A} = \begin{pmatrix} \mathbf{R} & \mathbf{I}_2 \\ \mathbf{O} & \mathbf{R} \end{pmatrix}$  with  $\mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ , so  $x\mathbf{I} - \mathbf{A}$  is block upper triangular and  $p_{\mathbf{A}} = (\det(x\mathbf{I}_2 - \mathbf{R}))^2 = (x^2+1)^2$  (Theorem 7.8.4), as stated.

*The minimal polynomial.* By Theorem 9.5.1 and Lemma 9.5.2,  $m_{\mathbf{A}}$  is  $x^2+1$  or  $(x^2+1)^2$ . Compute  $\mathbf{A}^2$ : its columns are  $\mathbf{A}$  applied to the columns of  $\mathbf{A}$ . With  $\mathbf{c}_1, \dots, \mathbf{c}_4$  the columns of  $\mathbf{A}$ ,

$$\mathbf{A}\mathbf{c}_1 = \mathbf{A}(0, 1, 0, 0) = (-1, 0, 0, 0), \quad \mathbf{A}\mathbf{c}_2 = \mathbf{A}(-1, 0, 0, 0) = (0, -1, 0, 0),$$

$$\mathbf{A}\mathbf{c}_3 = \mathbf{A}(1, 0, 0, 1) = (0, 1, 0, 0) + (0, 1, -1, 0) = (0, 2, -1, 0),$$

$$\mathbf{A}\mathbf{c}_4 = \mathbf{A}(0, 1, -1, 0) = (-1, 0, 0, 0) + (-1, 0, 0, -1) = (-2, 0, 0, -1).$$

So

$$\mathbf{A}^2 = \begin{pmatrix} -1 & 0 & 0 & -2 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad \mathbf{A}^2 + \mathbf{I} = \begin{pmatrix} 0 & 0 & 0 & -2 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \neq \mathbf{O}.$$

Hence  $m_{\mathbf{A}} = (x^2 + 1)^2 = p_{\mathbf{A}}$ , and  $\mathbf{A}$  is cyclic (Theorem 10.5.6). So  $r = 1$ , the single invariant factor is  $d_1 = (x^2 + 1)^2 = x^4 + 2x^2 + 1$ , and the rational canonical form is

$$\mathbf{C}(d_1) = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The elementary divisor form (Theorem 10.6.7) is the same matrix, since  $x^2 + 1$  is irreducible over  $\mathbb{Q}$  and  $d_1$  is already a prime power.

*A cyclic vector.* By Theorem 10.5.6 we need  $\mathbf{v}$  with  $m_{\mathbf{A},\mathbf{v}} = m_{\mathbf{A}}$ ; since  $\mathbf{A}^2 + \mathbf{I} \neq \mathbf{O}$ , pick  $\mathbf{v}$  with  $(\mathbf{A}^2 + \mathbf{I})\mathbf{v} \neq \mathbf{0}$ . The third column of  $\mathbf{A}^2 + \mathbf{I}$  is nonzero, so  $\mathbf{v} = \mathbf{e}_3$  works: then  $(\mathbf{A}^2 + \mathbf{I})\mathbf{v} \neq \mathbf{0}$ , so  $m_{\mathbf{A},\mathbf{v}} \nmid x^2 + 1$ , and as  $m_{\mathbf{A},\mathbf{v}}$  divides  $(x^2 + 1)^2$  it must equal  $(x^2 + 1)^2$  (Lemma 9.5.2). Its chain is

$$\begin{aligned} \mathbf{v} &= (0, 0, 1, 0), & \mathbf{A}\mathbf{v} &= (1, 0, 0, 1), & \mathbf{A}^2\mathbf{v} &= (0, 2, -1, 0), \\ \mathbf{A}^3\mathbf{v} &= \mathbf{A}(0, 2, -1, 0) = (-2, 0, 0, 0) + (-1, 0, 0, -1) = (-3, 0, 0, -1), \end{aligned}$$

reading columns of  $\mathbf{A}$  as before. Hence

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & -3 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}, \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{C}((x^2 + 1)^2).$$

*Invertibility of  $\mathbf{P}$ .* Expanding  $\det \mathbf{P}$  along the third row, whose entries are  $(1, 0, -1, 0)$  (Theorem 7.6.1), and then each  $3 \times 3$  determinant along its own first column,

$$\begin{aligned} \det \mathbf{P} &= 1 \cdot \det \begin{pmatrix} 1 & 0 & -3 \\ 0 & 2 & 0 \\ 1 & 0 & -1 \end{pmatrix} + (-1)(-1)^{3+3} \det \begin{pmatrix} 0 & 1 & -3 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} \\ &= 1 \cdot (2(-1 + 3)) - 0 = 4 \neq 0, \end{aligned}$$

the second determinant vanishing because of a zero row, and the first being  $2 \det \begin{pmatrix} 1 & -3 \\ 1 & -1 \end{pmatrix} = 2 \cdot 2 = 4$  by expansion along its second row.

*The check.* The identity  $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{C}(d_1)$  reads column by column as  $\mathbf{A}\mathbf{v} = \mathbf{A}\mathbf{v}$ ,  $\mathbf{A}(\mathbf{A}\mathbf{v}) = \mathbf{A}^2\mathbf{v}$ ,  $\mathbf{A}(\mathbf{A}^2\mathbf{v}) = \mathbf{A}^3\mathbf{v}$  — true by construction — and

$$\mathbf{A}(\mathbf{A}^3\mathbf{v}) = -\mathbf{v} - 2\mathbf{A}^2\mathbf{v},$$

which is the annihilator relation  $\mathbf{A}^4 + 2\mathbf{A}^2 + \mathbf{I} = \mathbf{O}$  applied to  $\mathbf{v}$ . Verifying it directly:  $\mathbf{A}(-3, 0, 0, -1) = -3(0, 1, 0, 0) - (0, 1, -1, 0) = (0, -4, 1, 0)$ , while

$$-\mathbf{v} - 2\mathbf{A}^2\mathbf{v} = -(0, 0, 1, 0) - 2(0, 2, -1, 0) = (0, -4, 2 - 1, 0) = (0, -4, 1, 0). \checkmark$$

## 10.6.8 Exercises

### 10.6.8.1 A. Check your understanding

#### Exercise. A1

- State the Cyclic Decomposition Theorem, including the divisibility chain and what  $d_r$  and  $\prod_j d_j$  are.
- Define the invariant factors and the elementary divisors of  $T$ .
- What is the rational canonical form of  $T$ , and over which fields does it exist?
- True or false: two matrices in  $M_n(\mathbb{Q})$  that are similar over  $\mathbb{C}$  are similar over  $\mathbb{Q}$ . Justify your answer.
- An operator on a 5-dimensional  $\mathbb{Q}$ -space has  $p_T = (x - 3)^5$  and  $m_T = (x - 3)^2$ . What are its elementary divisors and invariant factors?
- True or false:  $r = 1$  if and only if  $T$  is cyclic. Justify your answer.

*Solution.*

- For  $V \neq \{0\}$  finite-dimensional and  $T \in \mathcal{L}(V)$ , there are  $v_1, \dots, v_r$  with  $V = Z(v_1; T) \oplus \dots \oplus Z(v_r; T)$  and  $m_{T, v_j} = d_j$ , where  $d_1 \mid d_2 \mid \dots \mid d_r$  are monic and nonconstant,  $d_r = m_T$ , and  $d_1 \cdots d_r = p_T$  ([Theorem 10.6.2](#)).
- The invariant factors are the polynomials  $d_1 \mid \dots \mid d_r$  of that theorem; the elementary divisors are the prime powers occurring in their factorizations, listed with repetition ([Definition 10.6.1](#)). Both are determined by  $T$  ([Corollary 10.6.4](#)).
- The matrix  $\mathbf{C}(d_1) \oplus \dots \oplus \mathbf{C}(d_r)$ , which is  $[T]_{\mathcal{B}}^{\mathcal{B}}$  for a suitable basis; it exists over **every** field ([Theorem 10.6.5](#)).
- True, by [Corollary 10.6.10](#) with  $K = \mathbb{Q}$  and  $L = \mathbb{C}$ .
- Elementary divisors are powers of  $x - 3$  with largest exponent 2 and exponents summing to 5: so  $(x - 3)^2, (x - 3)^2, x - 3$ . The exponents, in decreasing order, are 2, 2, 1; numbering the columns backwards gives  $d_3 = (x - 3)^2 = m_T$ ,  $d_2 = (x - 3)^2$ , and  $d_1 = x - 3$ .
- True.  $r = 1$  means  $V = Z(v_1; T)$ , that is,  $v_1$  is a cyclic vector; conversely if  $v$  is cyclic, then  $d_1 = m_{T, v} = m_T = p_T$  is the only invariant factor, by Theorem 10.5.6 and the uniqueness in [Corollary 10.6.4](#).

### 10.6.8.2 B. Practice

#### Exercise. B1: Invariant factors from given data

In each case  $T$  is an operator on a  $\mathbb{Q}$ -vector space with the stated characteristic and minimal polynomials. Find the elementary divisors, the invariant factors, and the rational canonical form ([Theorem 10.6.5](#)).

- (a)  $p_T = (x - 2)^4$ ,  $m_T = (x - 2)^2$ .  
 (b)  $p_T = (x^2 - 2)^3$ ,  $m_T = (x^2 - 2)^2$ .  
 (c)  $p_T = (x - 1)^2(x + 1)^2$ ,  $m_T = (x - 1)(x + 1)$ .

*Solution.*

- (a) Powers of  $x - 2$  with largest exponent 2 and exponents summing to 4: either 2, 2 or 2, 1, 1. So the data do **not** determine the answer; both

$$\text{elementary divisors } (x - 2)^2, (x - 2)^2 \quad \text{and} \quad (x - 2)^2, x - 2, x - 2$$

are possible, with invariant factors  $((x - 2)^2, (x - 2)^2)$  and  $(x - 2, x - 2, (x - 2)^2)$  and rational forms

$$\mathbf{C}((x - 2)^2) \oplus \mathbf{C}((x - 2)^2), \quad (2) \oplus (2) \oplus \mathbf{C}((x - 2)^2),$$

where  $\mathbf{C}((x - 2)^2) = \mathbf{C}(x^2 - 4x + 4) = \begin{pmatrix} 0 & -4 \\ 1 & 4 \end{pmatrix}$ . (The extra datum that decides is  $\dim \ker(T - 2 \text{ id})$ , which is 2 in the first case and 3 in the second, by [Proposition 10.6.3](#).)

- (b)  $x^2 - 2$  is irreducible over  $\mathbb{Q}$ , since  $\sqrt{2} \notin \mathbb{Q}$  (Theorem 0.2.3, Theorem 6.3.1 (b)). The exponents are powers of  $x^2 - 2$  with largest 2 summing to 3: only 2, 1. So the elementary divisors are  $(x^2 - 2)^2$  and  $x^2 - 2$ , the invariant factors are  $d_1 = x^2 - 2$  and  $d_2 = (x^2 - 2)^2 = x^4 - 4x^2 + 4$ , and the rational form is

$$\mathbf{C}(x^2 - 2) \oplus \mathbf{C}((x^2 - 2)^2) = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & 0 & -4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

of size  $2 + 4 = 6 = \deg p_T$ .

- (c) Both primes have largest exponent 1 and total contribution 2, so the elementary divisors are  $x - 1, x - 1, x + 1, x + 1$ . Arranging by prime,  $x - 1 : 1, 1$  and  $x + 1 : 1, 1$ , so  $r = 2$  and  $d_2 = (x - 1)(x + 1) = m_T$ ,  $d_1 = (x - 1)(x + 1)$ . The rational form is

$$\mathbf{C}(x^2 - 1) \oplus \mathbf{C}(x^2 - 1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

(This operator is diagonalizable, with  $\text{diag}(1, 1, -1, -1)$  as its Jordan form; the rational form is a different matrix representing the same operator.)

### Exercise. B2: Rational forms of matrices

For each matrix, find  $p_A$ ,  $m_A$ , the invariant factors, the rational canonical form ([Theorem 10.6.5](#)), and the elementary divisor form ([Theorem 10.6.7](#)).

(a)  $A_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \in M_3(\mathbb{Q})$ .

(b)  $A_2 = \text{diag}(1, 1, 2) \in M_3(\mathbb{Q})$ .

$$(c) \mathbf{A}_3 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in M_4(\mathbb{Q}).$$

*Solution.*

- (a) Reading the columns,  $\mathbf{A}_1 \mathbf{e}_1 = \mathbf{e}_3$ ,  $\mathbf{A}_1 \mathbf{e}_3 = \mathbf{e}_2$ , and  $\mathbf{A}_1 \mathbf{e}_2 = \mathbf{e}_1$ , so  $\mathbf{A}_1^3 = \mathbf{I}$  and the polynomial  $x^3 - 1$  annihilates  $\mathbf{A}_1$ . Also  $\mathbf{A}_1^2 \mathbf{e}_1 = \mathbf{A}_1 \mathbf{e}_3 = \mathbf{e}_2$ , so  $(\mathbf{e}_1, \mathbf{A}_1 \mathbf{e}_1, \mathbf{A}_1^2 \mathbf{e}_1) = (\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2)$  is a basis of  $\mathbb{Q}^3$  and  $\mathbf{e}_1$  is a cyclic vector. Hence  $\deg m_{\mathbf{A}_1} = 3$  (Theorem 10.5.6), and since  $m_{\mathbf{A}_1}$  is monic of degree 3 dividing  $x^3 - 1$ , we get  $m_{\mathbf{A}_1} = p_{\mathbf{A}_1} = x^3 - 1$ . The single invariant factor is  $d_1 = x^3 - 1$ , and the rational canonical form is

$$\mathbf{C}(x^3 - 1) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Over  $\mathbb{Q}$ ,  $x^3 - 1 = (x - 1)(x^2 + x + 1)$  with both factors irreducible (the quadratic has no rational root). So the elementary divisors are  $x - 1$  and  $x^2 + x + 1$ , and the elementary divisor form is

$$(1) \oplus \mathbf{C}(x^2 + x + 1) = (1) \oplus \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}.$$

- (b)  $p_{\mathbf{A}_2} = (x - 1)^2(x - 2)$  and  $m_{\mathbf{A}_2} = (x - 1)(x - 2)$  (Proposition 9.5.6). Elementary divisors:  $x - 1, x - 1, x - 2$ . Invariant factors:  $d_2 = (x - 1)(x - 2) = m_{\mathbf{A}_2}$  and  $d_1 = x - 1$ . Rational form

$$(1) \oplus \mathbf{C}((x - 1)(x - 2)) = (1) \oplus \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix},$$

elementary divisor form  $(1) \oplus (1) \oplus (2) = \text{diag}(1, 1, 2) = \mathbf{A}_2$  itself.

- (c)  $\mathbf{A}_3 = \mathbf{R} \oplus \mathbf{R}$  with  $\mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ , and  $p_{\mathbf{R}} = x^2 + 1$ ,  $m_{\mathbf{R}} = x^2 + 1$ , so  $p_{\mathbf{A}_3} = (x^2 + 1)^2$  (Theorem 7.8.4 applied to  $x\mathbf{I} - \mathbf{A}_3$ , which is block diagonal over  $\mathbb{Q}[x]$ ) and  $m_{\mathbf{A}_3} = \text{lcm}(x^2 + 1, x^2 + 1) = x^2 + 1$  (Proposition 9.5.6). Since  $x^2 + 1$  is irreducible over  $\mathbb{Q}$ , the elementary divisors are  $x^2 + 1$  and  $x^2 + 1$ ;  $r = 2$  and  $d_1 = d_2 = x^2 + 1$ . Both forms are

$$\mathbf{C}(x^2 + 1) \oplus \mathbf{C}(x^2 + 1) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \mathbf{A}_3.$$

Compare [Example \(The Rational Canonical Form of a Matrix over the Rationals\)](#): that matrix has the same characteristic polynomial but a different minimal polynomial, so the two are **not** similar.

### Exercise. B3: Counting similarity classes over a finite field

Work over  $F = \mathbb{F}_2$ . For each polynomial  $p$ , find the number of similarity classes of matrices  $\mathbf{A} \in M_3(\mathbb{F}_2)$  with  $p_{\mathbf{A}} = p$ , and list the invariant factors of each class.

- (a)  $p = (x + 1)^3$ .  
 (b)  $p = x(x + 1)^2$ .

(c)  $p = x^3 + x + 1$ .

*Hint: by Corollary 10.6.6, a class is a way of writing  $p = d_1 \cdots d_r$  with  $d_1 \mid \cdots \mid d_r$ .*

*Solution.*

- (a) The only prime involved is  $x + 1$ , so the elementary divisors are powers  $(x + 1)^a$  with exponents forming a partition of 3. The three partitions (3), (2, 1), (1, 1, 1) give

$$((x + 1)^3), \quad (x + 1, (x + 1)^2), \quad (x + 1, x + 1, x + 1)$$

as lists of invariant factors. So there are 3 classes, with representatives  $\mathbf{C}((x + 1)^3)$ ,  $(1) \oplus \mathbf{C}((x + 1)^2)$ , and  $\mathbf{I}_3$  (note  $-1 = 1$  in  $\mathbb{F}_2$ ).

- (b) Now two primes,  $x$  with total exponent 1 and  $x + 1$  with total exponent 2. For  $x$  the only option is the single elementary divisor  $x$ ; for  $x + 1$  the options are  $(x + 1)^2$  or  $x + 1, x + 1$ . So there are 2 classes, with invariant factors

$$(x(x + 1)^2) \quad \text{and} \quad (x + 1, x(x + 1)).$$

In the first case  $r = 1$  and the matrix is cyclic; in the second  $r = 2$ .

- (c)  $x^3 + x + 1$  has no root in  $\mathbb{F}_2$  (its values at 0 and 1 are 1 and 1), so it is irreducible (Theorem 6.3.1 (b)). A prime of degree 3 in a characteristic polynomial of degree 3 can occur only with exponent 1, so the only possible list of elementary divisors is  $(x^3 + x + 1)$  itself: 1 class, represented by  $\mathbf{C}(x^3 + x + 1)$ .

### 10.6.8.3 C. Going deeper

#### Exercise. C1: Which matrices are companion matrices

Let  $\mathbf{A} \in M_n(F)$  with  $n \geq 1$ .

- Prove that  $\mathbf{A}$  is similar to a companion matrix if and only if  $m_{\mathbf{A}} = p_{\mathbf{A}}$ , and that the companion matrix is then  $\mathbf{C}(p_{\mathbf{A}})$ .
- Deduce that  $\mathbf{A}$  is similar to a companion matrix if and only if its rational canonical form has exactly one block.
- Give an example of a matrix in  $M_2(F)$  that is not similar to any companion matrix.

*Solution.*

- ( $\Leftarrow$ ) If  $m_{\mathbf{A}} = p_{\mathbf{A}}$ , then  $\mathbf{A}$  is cyclic (Theorem 10.5.6) and, in the basis generated by a cyclic vector, its matrix is  $\mathbf{C}(m_{\mathbf{A}}) = \mathbf{C}(p_{\mathbf{A}})$ ; so  $\mathbf{A} \sim \mathbf{C}(p_{\mathbf{A}})$  (Theorem 4.2.4).  
( $\Rightarrow$ ) Suppose  $\mathbf{A} \sim \mathbf{C}(q)$  for some monic  $q$ . By Theorem 10.5.3,  $m_{\mathbf{C}(q)} = p_{\mathbf{C}(q)} = q$ . Similar matrices share both polynomials (Theorem 7.10.3, Theorem 9.5.5), so  $m_{\mathbf{A}} = p_{\mathbf{A}} = q$ , and the companion matrix is  $\mathbf{C}(p_{\mathbf{A}})$ .
- By Theorem 10.6.5 the form has one block exactly when  $r = 1$ , that is, when  $d_1 = m_{\mathbf{A}} = p_{\mathbf{A}}$  (Theorem 10.6.2 (b), (c)). Now apply (a).
- $\mathbf{I}_2$ : here  $m = x - 1 \neq (x - 1)^2 = p$ . Concretely,  $\mathbf{I}_2$  is similar only to itself, and no companion matrix of a monic quadratic is  $\mathbf{I}_2$ , since the (2, 1) entry of  $\mathbf{C}(q)$  is  $1 \neq 0$ .

**Exercise. C2: All similarity classes of two by two matrices over the smallest field**

Work over  $F = \mathbb{F}_2$ , so that  $M_2(\mathbb{F}_2)$  has 16 elements.

- List the four monic polynomials of degree 2 over  $\mathbb{F}_2$  and factor each into irreducibles.
- For each, list the possible lists of invariant factors of a matrix  $\mathbf{A} \in M_2(\mathbb{F}_2)$  with that characteristic polynomial.
- Conclude the total number of similarity classes in  $M_2(\mathbb{F}_2)$ , and give a representative of each.
- Check your count against the orbit sizes:  $|\mathrm{GL}_2(\mathbb{F}_2)| = 6$ , and the class of  $\mathbf{A}$  has  $6/|\{\mathbf{P} : \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{A}\}|$  elements.

*Solution.*

- $x^2$ ;  $x^2 + 1 = (x + 1)^2$ ;  $x^2 + x = x(x + 1)$ ; and  $x^2 + x + 1$ , which has no root in  $\mathbb{F}_2$  and is therefore irreducible (Theorem 6.3.1 (b)).
- For  $p = x^2$ : elementary divisors  $x^2$  (invariant factors  $(x^2)$ ) or  $x, x$  (invariant factors  $(x, x)$ ) — two options. For  $p = (x + 1)^2$ : likewise two,  $((x + 1)^2)$  and  $(x + 1, x + 1)$ . For  $p = x(x + 1)$ : each prime occurs once, so the elementary divisors are  $x, x + 1$  and there is a single list of invariant factors,  $(x(x + 1))$ . For the irreducible  $p = x^2 + x + 1$ : one option,  $(x^2 + x + 1)$ .
- Total  $2 + 2 + 1 + 1 = 6$  classes, with representatives

$$\mathbf{C}(x^2) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{O}, \quad \mathbf{C}((x + 1)^2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\mathbf{I}_2, \quad \mathrm{diag}(0, 1), \quad \mathbf{C}(x^2 + x + 1) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

using  $-1 = 1$  in  $\mathbb{F}_2$ . (For  $(x + 1)^2 = x^2 + 1$  the coefficients are  $a_0 = 1, a_1 = 0$ , so  $\mathbf{C} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ , the coordinate swap.)

- The two scalar matrices  $\mathbf{O}$  and  $\mathbf{I}_2$  commute with everything, so their classes have  $6/6 = 1$  element each. For  $\mathrm{diag}(0, 1)$ , a matrix commutes with it exactly when it is diagonal (compare entries), and the invertible diagonal matrices over  $\mathbb{F}_2$  are only  $\mathbf{I}_2$ , so the class has 6 elements. For  $\mathbf{C}(x^2)$  and  $\mathbf{C}(x^2 + 1)$  the commutant inside  $\mathrm{GL}_2$  has 2 elements — by Exercise (C1: What commutes with a cyclic operator) the commutant of a cyclic matrix is  $\{a\mathbf{I} + b\mathbf{A}\}$ , which here has 4 elements, of which 2 are invertible — so each class has 3 elements. For  $\mathbf{C}(x^2 + x + 1)$ , whose order is 3 in  $\mathrm{GL}_2(\mathbb{F}_2)$ , the commutant is  $\{a\mathbf{I} + b\mathbf{A}\}$  again, with 3 invertible elements, so the class has 2 elements. Total:

$$1 + 1 + 6 + 3 + 3 + 2 = 16 = |M_2(\mathbb{F}_2)|. \checkmark$$

**Exercise. C3: A matrix over the rationals is similar to its transpose**

- Let  $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{Q})$  have the same Jordan form over  $\mathbb{C}$ . Prove that  $\mathbf{A}$  and  $\mathbf{B}$  are similar over  $\mathbb{Q}$ .
- Deduce that  $\mathbf{A} \sim \mathbf{A}^\top$  over  $\mathbb{Q}$  for every  $\mathbf{A} \in M_n(\mathbb{Q})$ .

- (c) Which step of (b) would fail for a general field  $K$  in place of  $\mathbb{Q}$ ?

*Hint for (b): Exercise (C1: A matrix is similar to its transpose).*

*Solution.*

- (a) A matrix in  $M_n(\mathbb{Q})$  is also a matrix in  $M_n(\mathbb{C})$ , and  $p_{\mathbf{A}}$  splits over  $\mathbb{C}$  (Corollary 6.5.5), so both have Jordan forms there. Having the same Jordan form, they are similar over  $\mathbb{C}$  (Corollary 10.3.3). By Corollary 10.6.10 with  $K = \mathbb{Q}$  and  $L = \mathbb{C}$ , they are similar over  $\mathbb{Q}$ .
- (b) Let  $\mathbf{A} \in M_n(\mathbb{Q})$ . Over  $\mathbb{C}$ ,  $p_{\mathbf{A}}$  splits, so Exercise (C1: A matrix is similar to its transpose) (c) gives  $\mathbf{A} \sim \mathbf{A}^{\top}$  over  $\mathbb{C}$ . Both matrices lie in  $M_n(\mathbb{Q})$ , so Corollary 10.6.10 gives  $\mathbf{A} \sim \mathbf{A}^{\top}$  over  $\mathbb{Q}$ .
- (c) The step that fails is the splitting: over a general field  $K$  there is no reason for  $p_{\mathbf{A}}$  to split, and we have not constructed a larger field where it does. (The conclusion is nevertheless true over every field; Section 7 proves it without any splitting, using the Smith normal form of  $x\mathbf{I} - \mathbf{A}$ .) Note that Corollary 10.6.10 itself needs no hypothesis on  $K$  — only the existence of a suitable  $L$  does.