

10.7 Polynomial Matrices and the Smith Normal Form

The rational canonical form of the previous section is a complete answer to the similarity problem, but the proof did not tell us how to *find* the invariant factors of a given matrix. This section supplies the missing algorithm, and it comes from an unexpected place: instead of working with $\mathbf{A}$, we work with the single polynomial matrix $x\mathbf{I} - \mathbf{A}$, and we reduce it by row and column operations whose scalars are allowed to be polynomials. The reduced form is diagonal, its nonconstant diagonal entries are the invariant factors, and two matrices are similar exactly when their reduced forms agree. On the way we finally pay off a promise from Chapter 4: every square matrix is similar to its transpose, over every field.

10.7.1 The space as a module over the polynomials

Everything in this chapter is built by applying **polynomials** in T to vectors: the T -annihilator of a vector, the cyclic subspace $Z(v; T)$, the kernels $\ker p(T)$. Name that structure once. For V a vector space over F and $T \in \mathcal{L}(V)$, the product

$$F[x] \times V \rightarrow V, \quad p \cdot v := p(T)v$$

satisfies, by Theorem 6.7.1, exactly the vector space axioms with F replaced by $F[x]$: it is additive in each factor, $(pq) \cdot v = p \cdot (q \cdot v)$, and $1 \cdot v = v$. A set with an addition and a product by the elements of a commutative ring R obeying these rules is an **R -module**, so the pair (V, T) is the same thing as an $F[x]$ -module that is finite-dimensional over the constants F . We develop no general theory of modules; the point is the dictionary:

$F[x]$ -module language	operator language
submodule of V	T -invariant subspace (Definition 9.1.1)
submodule generated by v , that is $F[x] \cdot v$	cyclic subspace $Z(v; T)$ (Definition 10.5.1)
the annihilator of v in $F[x]$	the T -annihilator of v
V is generated by one element	T has a cyclic vector

A subspace closed under multiplication by x is closed under multiplication by every polynomial, which is why the first two lines match.

Where does $x\mathbf{I} - \mathbf{A}$ come from here? Fix $\mathcal{B} = (b_1, \dots, b_n)$ with $\mathbf{A} = [T]_{\mathcal{B}}^{\mathcal{B}}$. As an $F[x]$ -module, V is generated by $b_1, \dots, b_n$, subject to the n relations $x \cdot b_j = \sum_i a_{ij} b_i$, whose coefficient vectors are the columns of $x\mathbf{I} - \mathbf{A}$. So $x\mathbf{I} - \mathbf{A}$ is the **matrix of relations**, and changing generators or replacing one relation by a combination of relations should not change V . That is the intuition behind the operations we now allow; the proofs below are matrix proofs and do not rely on it.

10.7.2 Row and column operations with polynomial multipliers

A matrix in $M_n(F[x])$ will be called a **polynomial matrix**. Over a field we reduced a matrix by three kinds of row operation (Definition 3.2.1). Over $F[x]$ we must be careful with one of

them: multiplying a row by x cannot be undone, because x has no inverse in $F[x]$. So the scaling operation keeps its scalar in F , while the operation that adds a multiple of one row to another may use any polynomial multiplier.

Definition 10.7.1. Elementary Operations over $F[x]$

Let $\mathbf{M} \in M_n(F[x])$. An **elementary row operation over $F[x]$** on $\mathbf{M}$ is one of:

- (E1) interchanging two rows;
- (E2) multiplying one row by a **nonzero scalar** $c \in F$;
- (E3) adding h times one row to a **different** row, where $h \in F[x]$ is **any** polynomial.

Elementary column operations over $F[x]$ are defined the same way with “column” in place of “row”.

In words: (E1) and (E3) are as over a field, except that (E3) now permits a polynomial multiplier; (E2) is restricted, because only the nonzero constants have inverses in $F[x]$. Each of the three is reversible by an operation of the same kind: swap back, scale by c^{-1} , add $-h$ times the row back.

Two polynomial matrices reachable from one another by such operations should count as the same. As over a field (Definition 4.3.1), the clean way to say this is with invertible matrices on both sides — but “invertible” must now mean invertible *inside* $M_n(F[x])$.

Proposition 10.7.1. Invertible Polynomial Matrices

Let $\mathbf{U} \in M_n(F[x])$. Then $\mathbf{U}$ has an inverse in $M_n(F[x])$ if and only if $\det \mathbf{U}$ is a **nonzero constant**.

Proof. ($\Rightarrow$) Suppose $\mathbf{U}\mathbf{W} = \mathbf{I}$ with $\mathbf{W} \in M_n(F[x])$. Taking determinants and using multiplicativity over the commutative ring $F[x]$ (Theorem 7.5.4 and the remark on commutative rings after it), $\det \mathbf{U} \cdot \det \mathbf{W} = 1$. By Theorem 0.8.3, $\deg(\det \mathbf{U}) + \deg(\det \mathbf{W}) = \deg 1 = 0$, and degrees of nonzero polynomials are ≥ 0 , so $\deg(\det \mathbf{U}) = 0$; that is, $\det \mathbf{U} \in F \setminus \{0\}$.

($\Leftarrow$) Suppose $\det \mathbf{U} = c \in F \setminus \{0\}$. The adjugate identity

$$\mathbf{U} \operatorname{adj} \mathbf{U} = \operatorname{adj} \mathbf{U} \mathbf{U} = (\det \mathbf{U}) \mathbf{I}$$

holds over any commutative ring (Theorem 7.6.2 and the remark on commutative rings following it), and $\operatorname{adj} \mathbf{U} \in M_n(F[x])$, since its entries are minors of $\mathbf{U}$. So $\mathbf{W} = c^{-1} \operatorname{adj} \mathbf{U} \in M_n(F[x])$ satisfies $\mathbf{U}\mathbf{W} = \mathbf{W}\mathbf{U} = \mathbf{I}$. ■

Definition 10.7.2. Equivalence over $F[x]$

Let $\mathbf{M}, \mathbf{M}' \in M_n(F[x])$. We say $\mathbf{M}$ and $\mathbf{M}'$ are **equivalent over $F[x]$** , written $\mathbf{M} \approx \mathbf{M}'$, if there are $\mathbf{U}, \mathbf{V} \in M_n(F[x])$, **each invertible in $M_n(F[x])$** , with

$$\mathbf{M}' = \mathbf{U}\mathbf{M}\mathbf{V}.$$

Proposition 10.7.2. Operations and Equivalence

Let $\mathbf{M}, \mathbf{M}' \in M_n(F[x])$.

- (a) $\approx$ is an equivalence relation on $M_n(F[x])$.
- (b) If $\mathbf{M}'$ is obtained from $\mathbf{M}$ by a finite sequence of elementary row and column operations over $F[x]$, then $\mathbf{M} \approx \mathbf{M}'$.

Proof.

- (a) Take $\mathbf{U} = \mathbf{V} = \mathbf{I}$ for reflexivity. If $\mathbf{M}' = \mathbf{U}\mathbf{M}\mathbf{V}$, then $\mathbf{M} = \mathbf{U}^{-1}\mathbf{M}'\mathbf{V}^{-1}$, and $\mathbf{U}^{-1}, \mathbf{V}^{-1}$ are invertible in $M_n(F[x])$; this gives symmetry. If $\mathbf{M}' = \mathbf{U}\mathbf{M}\mathbf{V}$ and $\mathbf{M}'' = \mathbf{U}'\mathbf{M}'\mathbf{V}'$, then $\mathbf{M}'' = (\mathbf{U}'\mathbf{U})\mathbf{M}(\mathbf{V}\mathbf{V}')$, and a product of invertible matrices is invertible, which gives transitivity.
- (b) Each elementary row operation over $F[x]$ is left multiplication by the matrix obtained by performing that operation on $\mathbf{I}_n$, exactly as over a field (Theorem 3.4.1, whose proof uses only the ring operations). Those three matrices are $\mathbf{P}_{ij}$, $\mathbf{D}_i(c)$, and $\mathbf{I} + h\mathbf{E}_{ij}$ with $i \neq j$, of determinants -1 , $c \neq 0$, and 1 respectively (Theorem 7.4.6 and Theorem 7.5.1, valid over a commutative ring), hence invertible in $M_n(F[x])$ by Proposition 10.7.1. Column operations are right multiplication by the same matrices. Now use (a).

■

Warning. Multiplying a row by a nonconstant polynomial is not an allowed operation. It changes the answer. The matrix $\mathbf{M} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ becomes $\begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix}$ after “multiplying row 1 by x ”, and the two are not equivalent: their determinants are 1 and x , and by Proposition 10.7.1 the first is invertible over $F[x]$ while the second is not. The rule to remember is that only the **units** of $F[x]$ — the nonzero constants — may be used as scaling factors.

10.7.3 The Smith normal form

Every polynomial matrix can be brought to a diagonal form whose diagonal entries divide one another in turn, and that form is unique.

The existence proof is Euclidean division, run on whichever nonzero entry has the smallest degree: if that entry fails to divide one of its neighbors in the same row or column, the remainder gives a nonzero entry of strictly smaller degree, and degrees cannot decrease forever.

Theorem 10.7.3. Smith Normal Form

Let F be a field, $n \geq 1$, and $\mathbf{M} \in M_n(F[x])$. Then:

- (a) **(Existence)** $\mathbf{M}$ is equivalent over $F[x]$, by a finite sequence of elementary row and column operations, to a diagonal matrix

$$\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$$

in which every d_i is **monic or zero** and $d_i \mid d_{i+1}$ for $i = 1, \dots, n-1$.

- (b) **(Uniqueness)** The list $(d_1, \dots, d_n)$ is determined by $\mathbf{M}$: if $\mathbf{M} \approx \text{diag}(d_1, \dots, d_n) \approx \text{diag}(e_1, \dots, e_n)$ with both lists as in (a), then $d_i = e_i$ for every i .

The matrix $\mathbf{D}$ is called the **Smith normal form** of $\mathbf{M}$, and $d_1, \dots, d_n$ its **invariant divisors**.

Since $0 \mid f$ holds only for $f = 0$, the divisibility chain forces the zero entries, if any, to be the last ones.

Idea. Step roadmap for existence. ① Among all matrices reachable from $\mathbf{M}$ by operations, pick one whose smallest nonzero entry degree is least, and move that entry to the corner. ② That corner entry f must divide every other entry of its row and column: otherwise divide with remainder and produce an entry of smaller degree, contradicting the choice in ①. ③ Clear the first row and column with f ; the same remainder trick, after adding a row to the first one, shows f divides every entry of what is left. ④ Induct on the size. The divisibility $d_1 \mid d_2$ survives the induction because operations replace entries by $F[x]$ -combinations of entries.

For uniqueness, find a quantity that can be read off both $\mathbf{M}$ and $\mathbf{D}$. The greatest common divisor of all $k \times k$ minors is such a quantity: minors of a product are $F[x]$ -combinations of minors of the factors (Cauchy–Binet), so the gcd cannot change under equivalence, and for a diagonal matrix with a divisibility chain it is simply $d_1 d_2 \cdots d_k$.

Proof.

- (a) If $\mathbf{M} = \mathbf{O}$, it is already of the stated form. So assume $\mathbf{M} \neq \mathbf{O}$.
For a nonzero $\mathbf{N} \in M_n(F[x])$ put

$$\delta(\mathbf{N}) := \min\{\deg n_{ij} : n_{ij} \neq 0\} \in \mathbb{N},$$

the least degree of a nonzero entry. Let $\mathcal{S}$ be the set of all nonzero matrices obtainable from $\mathbf{M}$ by finite sequences of elementary row and column operations over $F[x]$; it contains $\mathbf{M}$. The set $\{\delta(\mathbf{N}) : \mathbf{N} \in \mathcal{S}\}$ is a nonempty subset of $\mathbb{N}$, so it has a least element; **fix $\mathbf{N} \in \mathcal{S}$ with $\delta(\mathbf{N})$ least.**

Step 1. Put a least-degree entry in the corner, monic. Choose (i, j) with $n_{ij} \neq 0$ and $\deg n_{ij} = \delta(\mathbf{N})$. Swapping row i with row 1 and column j with column 1 (operations (E1)), and then scaling the new first row by the inverse of the leading coefficient of the corner entry (operation (E2)), we obtain a matrix in $\mathcal{S}$ whose $(1, 1)$ -entry f is **monic** of degree $\delta(\mathbf{N})$. These operations permute the entries and rescale some of them, so the new matrix still has least nonzero entry degree $\delta(\mathbf{N})$. Rename it $\mathbf{N}$.

Step 2. f divides every entry of the first row and the first column. Suppose $f \nmid n_{1j}$ for some $j \geq 2$. By Theorem 6.1.1, $n_{1j} = qf + r$ with $r \neq 0$ and $\deg r < \deg f$. Adding $-q$ times column 1 to column j (operation (E3)) produces a matrix $\mathbf{N}' \in \mathcal{S}$ whose $(1, j)$ -entry is $r \neq 0$ with $\deg r < \deg f = \delta(\mathbf{N})$. Then $\delta(\mathbf{N}') < \delta(\mathbf{N})$, contradicting the minimality of $\delta(\mathbf{N})$. Hence $f \mid n_{1j}$ for every j , and the same argument with rows in place of columns gives $f \mid n_{i1}$ for every i .

Step 3. Clear the first row and column. For $j \geq 2$ add $-(n_{1j}/f)$ times column 1 to column j ; the multipliers are polynomials by Step 2, and after these operations the first row is $(f, 0, \dots, 0)$ while the first column is unchanged. Then for $i \geq 2$ add $-(n_{i1}/f)$ times row 1 to row i ; since the first row is now $(f, 0, \dots, 0)$, this clears the

first column and changes nothing else. The result is

$$\begin{pmatrix} f & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{M}' \end{pmatrix} \in \mathcal{S}, \quad \mathbf{M}' \in M_{n-1}(F[x]),$$

and its least nonzero entry degree is still $\delta(\mathbf{N})$: it is $\geq \delta(\mathbf{N})$ by minimality, and f is an entry of degree $\delta(\mathbf{N})$.

Step 4. f divides every entry of $\mathbf{M}'$. Suppose $f \nmid (\mathbf{M}')_{kl}$ for some k, l . Add row $k+1$ to row 1. The new first row is $(f, (\mathbf{M}')_{k1}, \dots, (\mathbf{M}')_{k,n-1})$, so the $(1, l+1)$ -entry of the resulting matrix $\mathbf{N}'' \in \mathcal{S}$ is $(\mathbf{M}')_{kl}$, which is not divisible by f , while the $(1, 1)$ -entry is still f , monic of degree $\delta(\mathbf{N})$. By minimality $\delta(\mathbf{N}'') = \delta(\mathbf{N})$, so Step 2 applies to $\mathbf{N}''$ and says that f divides every entry of the first row of $\mathbf{N}''$ — a contradiction.

Step 5. Induct on n . We prove (a) by induction on n . For $n = 1$, scale the single entry to be monic (or leave it 0). For $n \geq 2$, Steps 1–4 give a sequence of operations taking $\mathbf{M}$ to $f \oplus \mathbf{M}'$ with f monic and f dividing every entry of $\mathbf{M}'$. By the induction hypothesis there are operations taking $\mathbf{M}'$ to $\mathbf{D}' = \text{diag}(d_2, \dots, d_n)$ with every d_i monic or zero and $d_2 \mid \dots \mid d_n$; performing the corresponding operations on rows and columns $2, \dots, n$ of $f \oplus \mathbf{M}'$ takes it to $f \oplus \mathbf{D}' = \text{diag}(f, d_2, \dots, d_n)$.

It remains to see that $f \mid d_2$. By Proposition 10.7.2, $\mathbf{D}' = \mathbf{U}'\mathbf{M}'\mathbf{V}'$ for some $\mathbf{U}', \mathbf{V}' \in M_{n-1}(F[x])$, so every entry of $\mathbf{D}'$ is a sum of products $u'_{ik}(\mathbf{M}')_{kl}v'_{lj}$ (Definition 0.9.9) and is therefore divisible by f , since every $(\mathbf{M}')_{kl}$ is. In particular $f \mid d_2$. Setting $d_1 = f$ finishes the induction.

- (b) Call a set $I \subseteq [n]$ with $|I| = k$ a **k -set**. For $\mathbf{M} \in M_n(F[x])$ and $1 \leq k \leq n$ define the k -th **determinantal divisor**

$$\Delta_k(\mathbf{M}) := \text{the monic generator of the ideal } \langle \det \mathbf{M}_{I,J} : I, J \text{ } k\text{-sets} \rangle \subseteq F[x],$$

and $\Delta_k(\mathbf{M}) := 0$ if all the $k \times k$ minors of $\mathbf{M}$ vanish. This is well defined by Proposition 6.2.1 and Theorem 6.2.2: the ideal generated by the finitely many $k \times k$ minors is $\langle \Delta_k(\mathbf{M}) \rangle$ for exactly one monic $\Delta_k(\mathbf{M})$ unless the ideal is $\{0\}$. Equivalently, $\Delta_k(\mathbf{M})$ is the monic greatest common divisor of the $k \times k$ minors: it divides each of them, and every common divisor of them divides it (Theorem 6.2.3, extended from two polynomials to finitely many by the same ideal argument).

Claim 1. $\mathbf{M} \approx \mathbf{M}'$ implies $\Delta_k(\mathbf{M}) = \Delta_k(\mathbf{M}')$ for every k .

Proof. Let $\mathbf{M}' = \mathbf{U}\mathbf{M}\mathbf{V}$ with $\mathbf{U}, \mathbf{V}$ invertible over $F[x]$, and let I, J be k -sets. The submatrix of a product on rows I and columns J is the product of the submatrix of the left factor on rows I and all columns with the submatrix of the right factor on all rows and columns J (Definition 0.9.9, read off entry by entry). Applying this twice and then Theorem 7.7.2 over the commutative ring $F[x]$ — legitimate by the remark after that theorem — to the $k \times n$ times $n \times k$ products involved,

$$\det \mathbf{M}'_{I,J} = \sum_S \det \mathbf{U}_{I,S} \det(\mathbf{M}\mathbf{V})_{S,J} = \sum_S \sum_R \det \mathbf{U}_{I,S} \det \mathbf{M}_{S,R} \det \mathbf{V}_{R,J},$$

with S and R running over the k -sets. So every $k \times k$ minor of $\mathbf{M}'$ is an $F[x]$ -combination of $k \times k$ minors of $\mathbf{M}$, hence lies in $\langle \det \mathbf{M}_{S,R} \rangle$. Therefore $\Delta_k(\mathbf{M})$ divides every $k \times k$ minor of $\mathbf{M}'$, and so $\Delta_k(\mathbf{M}) \mid \Delta_k(\mathbf{M}')$. Exchanging the roles of $\mathbf{M}$ and $\mathbf{M}'$ using $\mathbf{M} = \mathbf{U}^{-1}\mathbf{M}'\mathbf{V}^{-1}$ gives $\Delta_k(\mathbf{M}') \mid \Delta_k(\mathbf{M})$. Two monic polynomials dividing each other are equal (Proposition 6.1.2), and if one is 0 so is the other. ■

Claim 2. If $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ with $d_1 \mid d_2 \mid \dots \mid d_n$, then $\Delta_k(\mathbf{D}) = d_1 d_2 \cdots d_k$.

Proof. Let I, J be k -sets. If $I \neq J$, pick $i \in I \setminus J$; row i of $\mathbf{D}$ has its only possibly nonzero entry in column $i \notin J$, so the submatrix $\mathbf{D}_{I,J}$ has a zero row and $\det \mathbf{D}_{I,J} = 0$. If $I = J = \{i_1 < \dots < i_k\}$, then $\mathbf{D}_{I,I} = \text{diag}(d_{i_1}, \dots, d_{i_k})$ and $\det \mathbf{D}_{I,I} = d_{i_1} \cdots d_{i_k}$ (Theorem 7.4.6, valid over the commutative ring $F[x]$ by the remark after it). Since $i_j \geq j$, the chain gives $d_j \mid d_{i_j}$ for each j , hence

$$d_1 d_2 \cdots d_k \mid d_{i_1} d_{i_2} \cdots d_{i_k}.$$

So $d_1 \cdots d_k$ divides every $k \times k$ minor of $\mathbf{D}$, and it is itself the minor for $I = J = \{1, \dots, k\}$. Hence the ideal generated by the $k \times k$ minors is $\langle d_1 \cdots d_k \rangle$. If $d_1 \cdots d_k \neq 0$ it is monic, being a product of monic polynomials, and equals $\Delta_k(\mathbf{D})$. If $d_1 \cdots d_k = 0$, say $d_m = 0$ with $m \leq k$, then for every k -set I we have $i_k \geq k \geq m$, so $d_m \mid d_{i_k}$ forces $d_{i_k} = 0$ and the minor vanishes; thus $\Delta_k(\mathbf{D}) = 0 = d_1 \cdots d_k$. ■

Now let $\mathbf{M} \approx \text{diag}(d_1, \dots, d_n)$ with the d_i as in (a). By Claims 1 and 2,

$$\Delta_k(\mathbf{M}) = d_1 d_2 \cdots d_k \quad (k = 1, \dots, n),$$

a quantity depending only on $\mathbf{M}$. Reading this backwards recovers the d_i : first $d_1 = \Delta_1(\mathbf{M})$; and for $k \geq 2$, if $\Delta_{k-1}(\mathbf{M}) \neq 0$ then $d_k = \Delta_k(\mathbf{M})/\Delta_{k-1}(\mathbf{M})$, while if $\Delta_{k-1}(\mathbf{M}) = 0$ then $d_1 \cdots d_{k-1} = 0$, so some $d_m = 0$ with $m \leq k-1$, and the chain forces $d_k = 0$. Hence any second list $(e_1, \dots, e_n)$ as in (a) satisfies the same recipe and equals $(d_1, \dots, d_n)$. This proves the theorem. ■

The proof is an algorithm, and it is the one to use by hand: hunt for the entry of least degree, divide, and repeat.

Algorithm. **Smith normal form of $\mathbf{M} \in M_n(F[x])$.**

1. If $\mathbf{M} = \mathbf{O}$, stop. Otherwise find a nonzero entry of least degree and move it to position $(1, 1)$ by a row and a column swap; scale row 1 so that the entry is monic. Call it f .
2. For each other entry g in row 1 or column 1 with $f \nmid g$, replace g by its remainder on division by f (subtract a multiple of column 1, resp. row 1), move the remainder to the corner, and go back to step 1.
3. When f divides all of row 1 and column 1, clear them. If some entry g of the remaining block is not divisible by f , add its row to row 1 and go back to step 2.
4. Repeat on the remaining $(n-1) \times (n-1)$ block.
5. Check the answer: Δ_k must equal $d_1 \cdots d_k$. In particular $d_1 \cdots d_n$ is $\det \mathbf{M}$ rescaled by a nonzero constant to be monic, and d_1 is the monic gcd of the entries.

Quick check. Two polynomial matrices in $M_2(F[x])$ have determinants x^2 and x^2 , and in both the greatest common divisor of the four entries is 1. Are they equivalent over $F[x]$?

Solution. Yes. Here $\Delta_1 = 1$ (the gcd of the 1×1 minors, that is, of the entries) and Δ_2 is the determinant rescaled to be monic, so $\Delta_2 = x^2$ for both. By [Theorem 10.7.3](#) both have Smith normal form $\text{diag}(1, x^2)$, and since $\approx$ is an equivalence relation ([Proposition 10.7.2](#) (a)) they are equivalent to each other. For 2×2 matrices the pair (Δ_1, Δ_2) is therefore a complete invariant.

Example. A Smith Normal Form Computed by Hand

Let

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & -2 \\ 1 & 1 & -1 \end{pmatrix} \in M_3(\mathbb{Q}).$$

Find the Smith normal form of $x\mathbf{I} - \mathbf{A}$ over $\mathbb{Q}[x]$, naming every operation, and check the answer with determinantal divisors.

Solution. We have

$$x\mathbf{I} - \mathbf{A} = \begin{pmatrix} x-2 & -1 & 2 \\ -1 & x-2 & 2 \\ -1 & -1 & x+1 \end{pmatrix}.$$

Find the corner. The entries of least degree are the three entries equal to -1 . Swap columns 1 and 2, then multiply the first row by -1 :

$$\begin{pmatrix} -1 & x-2 & 2 \\ x-2 & -1 & 2 \\ -1 & -1 & x+1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -(x-2) & -2 \\ x-2 & -1 & 2 \\ -1 & -1 & x+1 \end{pmatrix}.$$

The corner is now 1, which divides everything, so no division step is needed.

Clear the first column. Add $-(x-2)$ times row 1 to row 2, and add row 1 to row 3. In row 2 this gives entries $-1 + (x-2)^2 = (x-1)(x-3)$ and $2 + 2(x-2) = 2(x-1)$; in row 3 it gives $-1 - (x-2) = -(x-1)$ and $x+1 - 2 = x-1$:

$$\begin{pmatrix} 1 & -(x-2) & -2 \\ 0 & (x-1)(x-3) & 2(x-1) \\ 0 & -(x-1) & x-1 \end{pmatrix}.$$

Clear the first row. Add $(x-2)$ times column 1 to column 2, and 2 times column 1 to column 3. Since the first column below the corner is zero, nothing else changes:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & (x-1)(x-3) & 2(x-1) \\ 0 & -(x-1) & x-1 \end{pmatrix} = \text{diag}(1) \oplus \begin{pmatrix} (x-1)(x-3) & 2(x-1) \\ -(x-1) & x-1 \end{pmatrix}.$$

Every entry of the 2×2 block is divisible by 1, so we move on to the block.

The block. Its entries of least degree are the three of degree 1. Swap rows 2 and 3 and scale

the new row 2 by -1 , putting the monic $x - 1$ in the corner of the block:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & x-1 & -(x-1) \\ 0 & (x-1)(x-3) & 2(x-1) \end{pmatrix}.$$

Now $x-1$ divides every entry of its row and column. Add column 2 to column 3, which clears the block's first row and turns the $(3, 3)$ -entry into $2(x-1) + (x-1)(x-3) = (x-1)(x-1)$:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & x-1 & 0 \\ 0 & (x-1)(x-3) & (x-1)^2 \end{pmatrix}.$$

Add $-(x-3)$ times row 2 to row 3 to clear the block's first column:

$$\mathbf{D} = \text{diag}(1, x-1, (x-1)^2),$$

and $1 \mid x-1 \mid (x-1)^2$, so by the uniqueness in [Theorem 10.7.3](#) this is **the** Smith normal form.

The check. The matrix $x\mathbf{I} - \mathbf{A}$ has an entry -1 , so $\Delta_1 = 1$. Its determinant is $p_{\mathbf{A}} = (x-1)^3$ — expanding along the first row,

$$\begin{aligned} & (x-2)[(x-2)(x+1)+2] + [-(x+1)+2] + 2[1+(x-2)] \\ &= (x-2)(x^2-x) + (1-x) + 2(x-1), \end{aligned}$$

which is $(x-2)x(x-1) + (x-1) = (x-1)(x^2-2x+1) = (x-1)^3$ — so $\Delta_3 = (x-1)^3$. For Δ_2 , the minor on rows $\{2, 3\}$ and columns $\{1, 2\}$ is $\det \begin{pmatrix} -1 & x-2 \\ -1 & -1 \end{pmatrix} = 1 + (x-2) = x-1$, so $\Delta_2 \mid x-1$; and every 2×2 minor is divisible by $x-1$, as Claim 2 predicts from $\mathbf{D}$. Hence $\Delta_2 = x-1$, and

$$(d_1, d_2, d_3) = \left(\Delta_1, \frac{\Delta_2}{\Delta_1}, \frac{\Delta_3}{\Delta_2} \right) = (1, x-1, (x-1)^2),$$

in agreement. As a sanity check on $\mathbf{A}$ itself: $\mathbf{A} - \mathbf{I}$ has rank 1, so $g_{\mathbf{A}}(1) = 2$ and the Jordan form of $\mathbf{A}$ is $\mathbf{J}_2(1) \oplus \mathbf{J}_1(1)$ (Corollary 10.3.2).

10.7.4 Similarity read off from a polynomial matrix

We can now state the theorem that makes all of this useful. It converts a question about $\mathbf{A}$ and $\mathbf{B}$ — is there an invertible $\mathbf{P} \in M_n(F)$ with $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$? — into a question about the two polynomial matrices $x\mathbf{I} - \mathbf{A}$ and $x\mathbf{I} - \mathbf{B}$, which the Smith algorithm answers by computation.

One direction is immediate. The other needs a tool: polynomial matrices can be divided by $x\mathbf{I} - \mathbf{A}$, with a **constant** remainder.

Lemma 10.7.4. Division of a Polynomial Matrix by $x\mathbf{I} - \mathbf{A}$

Let $\mathbf{A} \in M_n(F)$ and $\mathbf{M} \in M_n(F[x])$. Then there are $\mathbf{Q} \in M_n(F[x])$ and $\mathbf{R} \in M_n(F)$ with

$$\mathbf{M} = (x\mathbf{I} - \mathbf{A})\mathbf{Q} + \mathbf{R}.$$

Idea. Write $\mathbf{M}$ as a polynomial $\sum_k \mathbf{M}_k x^k$ with **constant matrix** coefficients. The scalar identity $x^k - a^k = (x - a)(x^{k-1} + \dots + a^{k-1})$ has a matrix version $x^k \mathbf{I} - \mathbf{A}^k = (x\mathbf{I} - \mathbf{A})\mathbf{S}_k$, in which the second factor is a polynomial in x and $\mathbf{A}$ and so commutes with $x\mathbf{I} - \mathbf{A}$. Replacing each $x^k \mathbf{I}$ by $(x^k \mathbf{I} - \mathbf{A}^k) + \mathbf{A}^k$ puts a factor $x\mathbf{I} - \mathbf{A}$ in front of everything except a constant matrix.

Proof. Collecting the coefficients of each power of x in the entries of $\mathbf{M}$, write $\mathbf{M} = \sum_{k=0}^m x^k \mathbf{M}_k$ with $\mathbf{M}_0, \dots, \mathbf{M}_m \in M_n(F)$; note that x^k is a scalar polynomial, so $x^k \mathbf{M}_k = \mathbf{M}_k x^k$. For $k \geq 1$ put

$$\mathbf{S}_k := \sum_{j=0}^{k-1} x^{k-1-j} \mathbf{A}^j \in M_n(F[x]), \quad \mathbf{S}_0 := \mathbf{O}.$$

Expanding the product and telescoping,

$$(x\mathbf{I} - \mathbf{A})\mathbf{S}_k = \sum_{j=0}^{k-1} x^{k-j} \mathbf{A}^j - \sum_{j=0}^{k-1} x^{k-1-j} \mathbf{A}^{j+1} = x^k \mathbf{I} - \mathbf{A}^k$$

for every $k \geq 0$. Therefore

$$\begin{aligned} \mathbf{M} &= \sum_{k=0}^m (x^k \mathbf{I}) \mathbf{M}_k = \sum_{k=0}^m (x^k \mathbf{I} - \mathbf{A}^k) \mathbf{M}_k + \sum_{k=0}^m \mathbf{A}^k \mathbf{M}_k \\ &= (x\mathbf{I} - \mathbf{A}) \underbrace{\sum_{k=0}^m \mathbf{S}_k \mathbf{M}_k}_{=\mathbf{Q}} + \underbrace{\sum_{k=0}^m \mathbf{A}^k \mathbf{M}_k}_{=\mathbf{R}}, \end{aligned}$$

with $\mathbf{Q} \in M_n(F[x])$ and $\mathbf{R} \in M_n(F)$, as claimed. ■

Theorem 10.7.5. Similarity as Equivalence of $x\mathbf{I} - \mathbf{A}$ and $x\mathbf{I} - \mathbf{B}$

Let F be a field, $n \geq 1$, and $\mathbf{A}, \mathbf{B} \in M_n(F)$. Then

$$\mathbf{A} \sim \mathbf{B} \quad \text{over } F \quad \Longleftrightarrow \quad x\mathbf{I} - \mathbf{A} \approx x\mathbf{I} - \mathbf{B} \quad \text{over } F[x].$$

Idea. $(\Rightarrow)$ is a one-line computation: a constant invertible $\mathbf{P}$ is also an invertible polynomial matrix.

$(\Leftarrow)$ is the substance, and the whole difficulty is that the matrices $\mathbf{U}, \mathbf{V}$ witnessing the equivalence have polynomial entries, while we must produce a **constant** $\mathbf{P}$. The trick is to divide $\mathbf{U}$ by $x\mathbf{I} - \mathbf{B}$ and keep only the constant remainder $\mathbf{P}$. Comparing degrees in the resulting identity shows that $\mathbf{P}$ alone already conjugates: $\mathbf{P}(x\mathbf{I} - \mathbf{A}) = (x\mathbf{I} - \mathbf{B})\mathbf{P}$, that is,

$\mathbf{PA} = \mathbf{BP}$. Running the same argument on the inverted relation produces a second constant $\mathbf{P}'$ with $\mathbf{P}'\mathbf{B} = \mathbf{AP}'$, and multiplying the two expressions for $\mathbf{V}^{-1}$ and $\mathbf{V}$ shows $\mathbf{PP}' = \mathbf{I}$, so $\mathbf{P}$ is invertible.

Proof. ($\Rightarrow$) Suppose $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$ with $\mathbf{P} \in M_n(F)$ invertible. Then, inside $M_n(F[x])$,

$$x\mathbf{I} - \mathbf{B} = \mathbf{P}^{-1}(x\mathbf{I})\mathbf{P} - \mathbf{P}^{-1}\mathbf{AP} = \mathbf{P}^{-1}(x\mathbf{I} - \mathbf{A})\mathbf{P},$$

and $\mathbf{P}^{-1}, \mathbf{P} \in M_n(F) \subseteq M_n(F[x])$ have nonzero constant determinants, hence are invertible over $F[x]$ (Proposition 10.7.1). So $x\mathbf{I} - \mathbf{A} \approx x\mathbf{I} - \mathbf{B}$.

($\Leftarrow$) Suppose

$$x\mathbf{I} - \mathbf{B} = \mathbf{U}(x\mathbf{I} - \mathbf{A})\mathbf{V} \quad (*)$$

with $\mathbf{U}, \mathbf{V} \in M_n(F[x])$ invertible over $F[x]$.

Step 1. Divide $\mathbf{U}$ by $x\mathbf{I} - \mathbf{B}$. By Lemma 10.7.4 there are $\mathbf{Q} \in M_n(F[x])$ and $\mathbf{P} \in M_n(F)$ with $\mathbf{U} = (x\mathbf{I} - \mathbf{B})\mathbf{Q} + \mathbf{P}$. Substituting into (*),

$$x\mathbf{I} - \mathbf{B} = (x\mathbf{I} - \mathbf{B})\mathbf{Q}(x\mathbf{I} - \mathbf{A})\mathbf{V} + \mathbf{P}(x\mathbf{I} - \mathbf{A})\mathbf{V},$$

so $\mathbf{P}(x\mathbf{I} - \mathbf{A})\mathbf{V} = (x\mathbf{I} - \mathbf{B})[\mathbf{I} - \mathbf{Q}(x\mathbf{I} - \mathbf{A})\mathbf{V}]$. Multiplying on the right by $\mathbf{V}^{-1}$ and putting

$$\mathbf{S} := \mathbf{V}^{-1} - \mathbf{Q}(x\mathbf{I} - \mathbf{A}) \in M_n(F[x]),$$

we get

$$\mathbf{P}(x\mathbf{I} - \mathbf{A}) = (x\mathbf{I} - \mathbf{B})\mathbf{S}. \quad (1)$$

Step 2. $\mathbf{S}$ is a constant matrix, and $\mathbf{PA} = \mathbf{BP}$. Suppose $\mathbf{S} \neq \mathbf{O}$ and write $\mathbf{S} = \sum_{k=0}^d x^k \mathbf{S}_k$ with $\mathbf{S}_k \in M_n(F)$ and $\mathbf{S}_d \neq \mathbf{O}$. The coefficient of x^{d+1} in $(x\mathbf{I} - \mathbf{B})\mathbf{S}$ is $\mathbf{S}_d \neq \mathbf{O}$, while the left-hand side of (1) is $x\mathbf{P} - \mathbf{PA}$, whose coefficients vanish beyond x^1 . Hence $d + 1 \leq 1$, that is $d = 0$. So in every case $\mathbf{S} \in M_n(F)$. Comparing the coefficients of x^1 in (1) gives $\mathbf{P} = \mathbf{S}$, and comparing the constant terms gives $-\mathbf{PA} = -\mathbf{BP}$. Therefore

$$\mathbf{PA} = \mathbf{BP}, \quad \text{equivalently} \quad \mathbf{P}(x\mathbf{I} - \mathbf{A}) = (x\mathbf{I} - \mathbf{B})\mathbf{P}, \quad (2)$$

the second form because $\mathbf{P}(x\mathbf{I} - \mathbf{A}) = x\mathbf{P} - \mathbf{PA} = x\mathbf{P} - \mathbf{BP} = (x\mathbf{I} - \mathbf{B})\mathbf{P}$. Moreover $\mathbf{S} = \mathbf{P}$ turns the definition of $\mathbf{S}$ into

$$\mathbf{V}^{-1} = \mathbf{P} + \mathbf{Q}(x\mathbf{I} - \mathbf{A}). \quad (3)$$

Step 3. The same argument, inverted. From (*), $x\mathbf{I} - \mathbf{A} = \mathbf{U}^{-1}(x\mathbf{I} - \mathbf{B})\mathbf{V}^{-1}$, which is (*) with $\mathbf{A}, \mathbf{B}$ interchanged and $\mathbf{U}, \mathbf{V}$ replaced by the invertible polynomial matrices $\mathbf{U}^{-1}, \mathbf{V}^{-1}$. So Steps 1 and 2 apply and give $\mathbf{Q}' \in M_n(F[x])$ and $\mathbf{P}' \in M_n(F)$ with

$$\mathbf{P}'(x\mathbf{I} - \mathbf{B}) = (x\mathbf{I} - \mathbf{A})\mathbf{P}' \quad \text{and} \quad \mathbf{V} = \mathbf{P}' + \mathbf{Q}'(x\mathbf{I} - \mathbf{B}). \quad (4)$$

Step 4. $\mathbf{P}$ is invertible. Multiplying (3) by (4),

$$\begin{aligned} \mathbf{I} &= \mathbf{V}^{-1}\mathbf{V} = [\mathbf{P} + \mathbf{Q}(x\mathbf{I} - \mathbf{A})][\mathbf{P}' + \mathbf{Q}'(x\mathbf{I} - \mathbf{B})] \\ &= \mathbf{PP}' + \mathbf{PQ}'(x\mathbf{I} - \mathbf{B}) + \mathbf{Q}(x\mathbf{I} - \mathbf{A})\mathbf{P}' \\ &\quad + \mathbf{Q}(x\mathbf{I} - \mathbf{A})\mathbf{Q}'(x\mathbf{I} - \mathbf{B}). \end{aligned}$$

By the first identity in (4), $(x\mathbf{I} - \mathbf{A})\mathbf{P}' = \mathbf{P}'(x\mathbf{I} - \mathbf{B})$, so the third summand is $\mathbf{Q}\mathbf{P}'(x\mathbf{I} - \mathbf{B})$ and

$$\mathbf{I} - \mathbf{P}\mathbf{P}' = \mathbf{Y}(x\mathbf{I} - \mathbf{B}), \quad \mathbf{Y} := \mathbf{P}\mathbf{Q}' + \mathbf{Q}\mathbf{P}' + \mathbf{Q}(x\mathbf{I} - \mathbf{A})\mathbf{Q}' \in M_n(F[x]).$$

If $\mathbf{Y} \neq \mathbf{O}$, write $\mathbf{Y} = \sum_{k=0}^e x^k \mathbf{Y}_k$ with $\mathbf{Y}_e \neq \mathbf{O}$; the coefficient of x^{e+1} on the right is $\mathbf{Y}_e \neq \mathbf{O}$, while the left-hand side is constant. Hence $\mathbf{Y} = \mathbf{O}$ and $\mathbf{P}\mathbf{P}' = \mathbf{I}$. By Theorem 3.4.5, $\mathbf{P}$ is invertible with $\mathbf{P}^{-1} = \mathbf{P}'$.

Step 5. Conclude. By (2), $\mathbf{P}\mathbf{A} = \mathbf{B}\mathbf{P}$, so $\mathbf{B} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$ and $\mathbf{A} \sim \mathbf{B}$. This proves the theorem. ■

Two by-products are worth recording. Whenever $x\mathbf{I} - \mathbf{B} = \mathbf{U}(x\mathbf{I} - \mathbf{A})\mathbf{V}$ with $\mathbf{U}, \mathbf{V}$ invertible over $F[x]$, taking determinants gives $p_{\mathbf{B}} = (\det \mathbf{U})(\det \mathbf{V}) p_{\mathbf{A}}$ with $\det \mathbf{U}, \det \mathbf{V}$ nonzero constants (Proposition 10.7.1); both $p_{\mathbf{A}}$ and $p_{\mathbf{B}}$ are monic of degree n , so

$$p_{\mathbf{A}} = p_{\mathbf{B}} \quad \text{and} \quad (\det \mathbf{U})(\det \mathbf{V}) = 1,$$

which recovers the invariance of the characteristic polynomial. And, by Theorem 10.7.3, the test is now decidable by a finite computation in $F[x]$: reduce $x\mathbf{I} - \mathbf{A}$ and $x\mathbf{I} - \mathbf{B}$ to Smith normal form and compare.

Corollary 10.7.6. The Smith Form of $x\mathbf{I} - \mathbf{A}$ Carries the Invariant Factors

Let $\mathbf{A} \in M_n(F)$ with invariant factors $d_1 \mid d_2 \mid \cdots \mid d_r$ as in Theorem 10.6.5. Then the Smith normal form of $x\mathbf{I} - \mathbf{A}$ is

$$\text{diag}(\underbrace{1, \dots, 1}_{n-r}, d_1, d_2, \dots, d_r).$$

In particular the invariant factors of $\mathbf{A}$ are exactly the nonconstant diagonal entries of the Smith normal form of $x\mathbf{I} - \mathbf{A}$, read from the top down, and two matrices in $M_n(F)$ are similar if and only if $x\mathbf{I} - \mathbf{A}$ and $x\mathbf{I} - \mathbf{B}$ have the same Smith normal form.

Proof. By Theorem 10.6.5, $\mathbf{A} \sim \mathbf{C}(d_1) \oplus \cdots \oplus \mathbf{C}(d_r)$ with d_i monic, $\deg d_i = k_i \geq 1$, and $\sum k_i = n$. By the ($\Rightarrow$) direction of Theorem 10.7.5 it is enough to find the Smith normal form of

$$x\mathbf{I}_n - (\mathbf{C}(d_1) \oplus \cdots \oplus \mathbf{C}(d_r)) = (x\mathbf{I}_{k_1} - \mathbf{C}(d_1)) \oplus \cdots \oplus (x\mathbf{I}_{k_r} - \mathbf{C}(d_r)),$$

the equality by Theorem 8.1.4 (a).

One block. Let d be monic of degree $k \geq 1$. By Definition 10.5.3, the only nonzero entries of $x\mathbf{I}_k - \mathbf{C}(d)$ are x on the diagonal (with $x + a_{k-1}$ in the corner), -1 in each position $(i+1, i)$, and the entries of the last column. Delete row 1 and column k . In the remaining $(k-1) \times (k-1)$ matrix, the entry in position (p, q) is the $(p+1, q)$ -entry of $x\mathbf{I}_k - \mathbf{C}(d)$: it is -1 when $q = p$, it is x when $q = p+1$, and it is 0 when $q < p$. So that matrix is upper triangular with every diagonal entry -1 , and its determinant is $(-1)^{k-1} \neq 0$ by Theorem 7.4.6, again over $F[x]$. Hence $\Delta_{k-1}(x\mathbf{I}_k - \mathbf{C}(d)) = 1$. By Claim 2 in the proof of Theorem 10.7.3, the determinantal divisors of any matrix are the partial products $d_1 \cdots d_j$ of its Smith diagonal, so $\Delta_j \mid \Delta_{k-1} = 1$ and therefore $\Delta_j = 1$ for every $j \leq k-1$. Also $\Delta_k = \det(x\mathbf{I}_k - \mathbf{C}(d)) = p_{\mathbf{C}(d)} = d$ by Theorem 10.5.3. By the uniqueness part of Theorem 10.7.3, the Smith normal form of $x\mathbf{I}_k - \mathbf{C}(d)$ is $\text{diag}(1, \dots, 1, d)$. (For $k = 1$ read this as $\text{diag}(d)$.)

Assembling. If $\mathbf{M}_i \approx \mathbf{M}'_i$ with witnesses $\mathbf{U}_i, \mathbf{V}_i$, then $\bigoplus_i \mathbf{M}_i \approx \bigoplus_i \mathbf{M}'_i$ with witnesses $\bigoplus_i \mathbf{U}_i$ and $\bigoplus_i \mathbf{V}_i$, which are invertible with inverses $\bigoplus_i \mathbf{U}_i^{-1}, \bigoplus_i \mathbf{V}_i^{-1}$ (Theorem 8.1.4 (a), (c)). So $x\mathbf{I}_n - \mathbf{A}$ is equivalent to the direct sum of the matrices $\text{diag}(1, \dots, 1, d_i)$, a diagonal matrix whose diagonal entries are $n - r$ ones and the polynomials $d_1, \dots, d_r$. Simultaneous row and column swaps — each a pair of elementary operations — reorder these entries at will, so $x\mathbf{I}_n - \mathbf{A}$ is equivalent to

$$\text{diag}(1, \dots, 1, d_1, d_2, \dots, d_r),$$

and this is already in the shape required by Theorem 10.7.3 (a), because $1 \mid d_1 \mid d_2 \mid \dots \mid d_r$ and all the entries are monic. By uniqueness it *is* the Smith normal form of $x\mathbf{I} - \mathbf{A}$. The final claim combines this with Theorem 10.7.5 and the uniqueness in Theorem 10.7.3. ■

So the determinantal divisors of $x\mathbf{I} - \mathbf{A}$ compute the invariant factors of $\mathbf{A}$ without any factorization of polynomials: Δ_k is a gcd, obtained from the Euclidean algorithm alone. Moreover $\Delta_n(x\mathbf{I} - \mathbf{A}) = p_{\mathbf{A}}$ and $\Delta_n/\Delta_{n-1} = d_r = m_{\mathbf{A}}$.

10.7.5 Every matrix is similar to its transpose

Chapter 4 asked whether $\mathbf{A}$ and $\mathbf{A}^\top$ are always similar and postponed the answer; Exercise (C1: A matrix is similar to its transpose) settled it whenever $p_{\mathbf{A}}$ splits. The Smith form removes that hypothesis, because transposing a matrix does not change its minors.

Corollary 10.7.7. A Matrix Is Similar to Its Transpose

Let F be any field and $\mathbf{A} \in M_n(F)$. Then $\mathbf{A} \sim \mathbf{A}^\top$.

Proof. Let $\mathbf{M} = x\mathbf{I} - \mathbf{A}$, so that $\mathbf{M}^\top = x\mathbf{I} - \mathbf{A}^\top$. For k -sets I, J we have $(\mathbf{M}^\top)_{I,J} = (\mathbf{M}_{J,I})^\top$ by Definition 0.9.10, so $\det(\mathbf{M}^\top)_{I,J} = \det \mathbf{M}_{J,I}$ by Theorem 7.4.5, valid over the commutative ring $F[x]$. Hence the $k \times k$ minors of $\mathbf{M}^\top$ are exactly the $k \times k$ minors of $\mathbf{M}$, listed in a different order, and

$$\Delta_k(x\mathbf{I} - \mathbf{A}^\top) = \Delta_k(x\mathbf{I} - \mathbf{A}) \quad (k = 1, \dots, n).$$

By Theorem 10.7.3, each of $x\mathbf{I} - \mathbf{A}$ and $x\mathbf{I} - \mathbf{A}^\top$ is equivalent over $F[x]$ to the diagonal matrix built from its own determinantal divisors, and these two diagonal matrices coincide. Since $\approx$ is an equivalence relation (Proposition 10.7.2 (a)), $x\mathbf{I} - \mathbf{A} \approx x\mathbf{I} - \mathbf{A}^\top$. By Theorem 10.7.5, $\mathbf{A} \sim \mathbf{A}^\top$. ■

Two sentences of consequence. An operator and its dual have the same invariants: by Theorem 5.4.2 the matrix of T' in the dual basis is $([T]_{\mathcal{B}}^{\mathcal{B}})^\top$, so T and T' share their characteristic polynomial, their minimal polynomial, and their invariant factors. And a matrix and its transpose always have the same Jordan form when one exists, which says that the *rows* of $\mathbf{A} - \lambda\mathbf{I}$ are exactly as informative as its columns.

Warning. **Similar is not the same** as “equivalent over $F[x]$ ” for the matrices themselves. The theorem compares $x\mathbf{I} - \mathbf{A}$ with $x\mathbf{I} - \mathbf{B}$, not $\mathbf{A}$ with $\mathbf{B}$. As polynomial matrices, any two invertible $\mathbf{A}, \mathbf{B} \in M_n(F)$ are equivalent over $F[x]$ — take $\mathbf{U} = \mathbf{A}^{-1}$, $\mathbf{V} = \mathbf{B}$ — whether or not they are similar. The whole content of [Theorem 10.7.5](#) is that inserting the variable x on the diagonal remembers the conjugation.

10.7.6 Exercises

10.7.6.1 A. Check your understanding

Exercise. A1

- List the three elementary operations over $F[x]$, and say which one differs from its counterpart over a field and why.
- State a criterion for a matrix $\mathbf{U} \in M_n(F[x])$ to be invertible in $M_n(F[x])$.
- State the Smith Normal Form theorem, including the uniqueness clause.
- Define the k -th determinantal divisor of a polynomial matrix, and say why it is unchanged by equivalence.
- True or false: if $\det(x\mathbf{I} - \mathbf{A}) = \det(x\mathbf{I} - \mathbf{B})$, then $x\mathbf{I} - \mathbf{A} \approx x\mathbf{I} - \mathbf{B}$. Justify your answer.
- How do you read $m_{\mathbf{A}}$ off the Smith normal form of $x\mathbf{I} - \mathbf{A}$?

Solution.

- Swap two rows; multiply a row by a nonzero **scalar** of F ; add a **polynomial** multiple of one row to a different row ([Definition 10.7.1](#)). The scaling operation is restricted, because only the nonzero constants are invertible in $F[x]$; multiplying a row by x is not reversible.
- $\mathbf{U}$ is invertible over $F[x]$ if and only if $\det \mathbf{U}$ is a nonzero constant ([Proposition 10.7.1](#)).
- Every $\mathbf{M} \in M_n(F[x])$ is equivalent over $F[x]$, by elementary operations, to a diagonal $\text{diag}(d_1, \dots, d_n)$ with each d_i monic or zero and $d_i \mid d_{i+1}$; this diagonal is unique ([Theorem 10.7.3](#)).
- $\Delta_k(\mathbf{M})$ is the monic generator of the ideal generated by all $k \times k$ minors of $\mathbf{M}$, equivalently their monic gcd. It is unchanged by equivalence because every $k \times k$ minor of $\mathbf{UMV}$ is an $F[x]$ -combination of $k \times k$ minors of $\mathbf{M}$, by the Cauchy–Binet formula (Theorem 7.7.2; Claim 1 in the proof of [Theorem 10.7.3](#)).
- False. Equality of determinants says only $\Delta_n(x\mathbf{I} - \mathbf{A}) = \Delta_n(x\mathbf{I} - \mathbf{B})$; the lower determinantal divisors can differ. For $\mathbf{A} = \mathbf{J}_2(0) \oplus \mathbf{J}_2(0)$ and $\mathbf{B} = \mathbf{J}_3(0) \oplus \mathbf{J}_1(0)$ in $M_4(F)$, both determinants are x^4 , but the Smith normal forms are $\text{diag}(1, 1, x^2, x^2)$ and $\text{diag}(1, 1, x, x^3)$. (Equivalently, the matrices are not similar, by Corollary 10.3.3.)
- $m_{\mathbf{A}}$ is the **last** diagonal entry of the Smith normal form of $x\mathbf{I} - \mathbf{A}$, that is $\Delta_n/\Delta_{n-1} = p_{\mathbf{A}}/\Delta_{n-1}$ ([Corollary 10.7.6](#), since the last diagonal entry is $d_r = m_{\mathbf{A}}$ in the notation there).

10.7.6.2 B. Practice

Exercise. B1: Smith normal forms

Find the Smith normal form over $\mathbb{Q}[x]$ of each of the following, naming the operations you use.

- (a) $\mathbf{M}_1 = \begin{pmatrix} x & x+1 \\ x-1 & x \end{pmatrix}.$
- (b) $\mathbf{M}_2 = \begin{pmatrix} x^2-1 & 0 \\ 0 & x^2-x \end{pmatrix}.$
- (c) $\mathbf{M}_3 = \begin{pmatrix} x & 0 & 1 \\ -1 & x & 0 \\ 0 & -1 & x \end{pmatrix}.$

Solution.

- (a) $\det \mathbf{M}_1 = x^2 - (x+1)(x-1) = 1$, so $\mathbf{M}_1$ is invertible over $\mathbb{Q}[x]$ ([Proposition 10.7.1](#)) and $\Delta_1 = \Delta_2 = 1$. The Smith normal form is $\text{diag}(1, 1) = \mathbf{I}_2$. (Concretely: subtract column 1 from column 2 to get $\begin{pmatrix} x & 1 \\ x-1 & 1 \end{pmatrix}$, then subtract row 1 from row 2, swap columns, and clear.)
- (b) Here $\Delta_1 = \gcd(x^2-1, x^2-x) = \gcd((x-1)(x+1), x(x-1)) = x-1$ and $\Delta_2 = \det \mathbf{M}_2 = x(x-1)^2(x+1)$. Reading the determinantal divisors backwards as in the proof of [Theorem 10.7.3](#), the Smith normal form is

$$\text{diag}\left(x-1, \frac{x(x-1)^2(x+1)}{x-1}\right) = \text{diag}(x-1, x(x-1)(x+1)),$$

and indeed $x-1$ divides $x(x-1)(x+1)$. By operations: add column 2 to column 1, giving first column (x^2-1, x^2-x) ; subtract row 1 from row 2 to get $(x^2-1, 1-x)$ in column 1; the entry $1-x$ has least degree, so swap it into the corner, scale by -1 , and continue.

- (c) The 2×2 minor on rows $\{2, 3\}$ and columns $\{1, 2\}$ is $\det \begin{pmatrix} -1 & x \\ 0 & -1 \end{pmatrix} = 1$, so $\Delta_2 = 1$, and then $\Delta_1 = 1$ as well (it divides Δ_2). Expanding $\det \mathbf{M}_3$ along the first row, $x(x^2-0) + 1 \cdot (1-0) = x^3+1$, so $\Delta_3 = x^3+1$ and the Smith normal form is

$\text{diag}(1, 1, x^3+1)$. Note $\mathbf{M}_3 = x\mathbf{I} - \mathbf{C}$ with $\mathbf{C} = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \mathbf{C}(x^3+1)$, and

[Corollary 10.7.6](#) predicts exactly this, since a companion matrix has the single invariant factor x^3+1 .

Exercise. B2: Invariant factors from determinantal divisors

For each matrix $\mathbf{A}$, compute $\Delta_1, \dots, \Delta_n$ of $x\mathbf{I} - \mathbf{A}$, write down the Smith normal form, and read off the invariant factors, $m_{\mathbf{A}}$, and the Jordan form.

- (a) $\mathbf{A} = \mathbf{J}_2(0) \oplus \mathbf{J}_2(0) \in M_4(\mathbb{Q})$.
- (b) $\mathbf{A} = \mathbf{J}_3(0) \oplus \mathbf{J}_1(0) \in M_4(\mathbb{Q})$.

$$(c) \mathbf{A} = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} \in M_2(\mathbb{Q}).$$

Solution. One remark makes the minors easy. If $\mathbf{M} = \mathbf{M}_1 \oplus \mathbf{M}_2$ with square blocks, and I, J are k -sets, split $I = I_1 \cup I_2$ and $J = J_1 \cup J_2$ according to the blocks. If $|I_1| \neq |J_1|$, then $\mathbf{M}_{I,J}$ has a zero row or a zero column and $\det \mathbf{M}_{I,J} = 0$; if $|I_1| = |J_1|$, then $\mathbf{M}_{I,J} = (\mathbf{M}_1)_{I_1,J_1} \oplus (\mathbf{M}_2)_{I_2,J_2}$ and the determinant is the product of the two, by Theorem 7.8.4, whose proof uses only sums and products of entries and so applies over $F[x]$ (the remark after that theorem). So **the minors of a block diagonal matrix are the products of matching-size minors of the blocks**. Here $x\mathbf{I} - \mathbf{A}$ is the direct sum of the blocks $x\mathbf{I}_k - \mathbf{J}_k(0)$, each upper triangular with x on the diagonal and -1 on the superdiagonal.

- (a) $x\mathbf{I} - \mathbf{A} = \mathbf{N} \oplus \mathbf{N}$ with $\mathbf{N} = \begin{pmatrix} x & -1 \\ 0 & x \end{pmatrix}$, whose minors are the entries $x, -1, 0$ for $k = 1$ and x^2 for $k = 2$. By the remark, the minors of $\mathbf{N} \oplus \mathbf{N}$ are $\{x, -1, 0\}$ for $k = 1$; the products of two 1×1 minors together with x^2 for $k = 2$; x^2 times a 1×1 minor for $k = 3$; and x^4 for $k = 4$. Hence

$$\Delta_1 = 1, \quad \Delta_2 = 1 \quad (\text{from } (-1)(-1) = 1), \quad \Delta_3 = x^2, \quad \Delta_4 = x^4,$$

so by Theorem 10.7.3 the Smith normal form is $\text{diag}(1, 1, x^2, x^2)$, the invariant factors are $x^2 \mid x^2$, $m_{\mathbf{A}} = x^2$ is the last of them, and the Jordan form is $\mathbf{J}_2(0) \oplus \mathbf{J}_2(0)$.

- (b) Now $x\mathbf{I} - \mathbf{A} = \mathbf{N}_3 \oplus (x)$ with $\mathbf{N}_3 = \begin{pmatrix} x & -1 & 0 \\ 0 & x & -1 \\ 0 & 0 & x \end{pmatrix}$. The minors of $\mathbf{N}_3$ are $\{x, -1, 0\}$ for $k = 1$; $\{x^2, -x, 1, 0\}$ for $k = 2$, the value 1 coming from rows $\{1, 2\}$ and columns $\{2, 3\}$; and x^3 for $k = 3$. By the remark, the $k \times k$ minors of $x\mathbf{I} - \mathbf{A}$ are those of $\mathbf{N}_3$ together with x times its $(k-1) \times (k-1)$ minors, so

$$\begin{aligned} \Delta_1 &= 1, \\ \Delta_2 &= \gcd(x^2, -x, 1, x \cdot x, -x) = 1, \\ \Delta_3 &= \gcd(x^3, x \cdot x^2, -x \cdot x, x \cdot 1) = x, \\ \Delta_4 &= x^4. \end{aligned}$$

The Smith normal form is $\text{diag}(1, 1, x, x^3)$: the invariant factors are $x \mid x^3$, $m_{\mathbf{A}} = x^3$ is the last of them, and the Jordan form is $\mathbf{J}_3(0) \oplus \mathbf{J}_1(0)$. Compared with (a) the determinants agree but Δ_3 does not, which is exactly why the two matrices are not similar.

- (c) $x\mathbf{I} - \mathbf{A} = \begin{pmatrix} x-3 & -1 \\ 0 & x-3 \end{pmatrix}$. Then $\Delta_1 = 1$ and $\Delta_2 = (x-3)^2$, so the Smith normal form is $\text{diag}(1, (x-3)^2)$, the only invariant factor is $(x-3)^2 = m_{\mathbf{A}} = p_{\mathbf{A}}$, and the Jordan form is $\mathbf{J}_2(3)$. Consistently, $\mathbf{A}$ is a companion matrix up to similarity, since $m_{\mathbf{A}} = p_{\mathbf{A}}$ (Theorem 10.5.6).

Exercise. B3: Deciding similarity

Determine whether each pair is similar over the stated field. Justify your answer using Theorem 10.7.5.

- (a) $\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ over $\mathbb{Q}$.
- (b) $\mathbf{A} = \begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & -2 \\ 1 & 1 & -1 \end{pmatrix}$ and $\mathbf{B} = \mathbf{J}_2(1) \oplus \mathbf{J}_1(1)$ over $\mathbb{Q}$.
- (c) $\mathbf{A} = \begin{pmatrix} 0 & -2 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & -1 \\ 2 & 0 \end{pmatrix}$ over $\mathbb{Q}$.

Solution.

- (a) Similar. Here $\mathbf{B} = \mathbf{A}^\top$, so [Corollary 10.7.7](#) applies. (Explicitly, $\Delta_1 = 1$ and $\Delta_2 = x^2 + 1$ for both, so both Smith normal forms are $\text{diag}(1, x^2 + 1)$.)
- (b) Similar. By [Example \(A Smith Normal Form Computed by Hand\)](#) the Smith normal form of $x\mathbf{I} - \mathbf{A}$ is $\text{diag}(1, x - 1, (x - 1)^2)$. For $\mathbf{B}$, $x\mathbf{I} - \mathbf{B} = \begin{pmatrix} x - 1 & -1 \\ 0 & x - 1 \end{pmatrix} \oplus (x - 1)$. Its entry -1 gives $\Delta_1 = 1$. By the block-minor rule in the solution to [Exercise \(B2: Invariant factors from determinantal divisors\)](#), the 2×2 minors are $(x - 1)^2$, from the 2×2 block, together with the products of a 1×1 minor of each block, that is $(x - 1)^2$, $-(x - 1)$, and 0 ; so $\Delta_2 = x - 1$. Finally $\Delta_3 = \det(x\mathbf{I} - \mathbf{B}) = (x - 1)^3$. The Smith normal forms agree, so $x\mathbf{I} - \mathbf{A} \approx x\mathbf{I} - \mathbf{B}$ and $\mathbf{A} \sim \mathbf{B}$.

- (c) Similar. Here

$$x\mathbf{I} - \mathbf{A} = \begin{pmatrix} x & 2 \\ -1 & x \end{pmatrix}, \quad x\mathbf{I} - \mathbf{B} = \begin{pmatrix} x & 1 \\ -2 & x \end{pmatrix},$$

and each has an entry that is a nonzero constant, so $\Delta_1 = 1$ in both cases; and $\Delta_2 = \det(x\mathbf{I} - \mathbf{A}) = x^2 + 2 = \det(x\mathbf{I} - \mathbf{B})$. So both Smith normal forms are $\text{diag}(1, x^2 + 2)$, hence $x\mathbf{I} - \mathbf{A} \approx x\mathbf{I} - \mathbf{B}$ and $\mathbf{A} \sim \mathbf{B}$ by [Theorem 10.7.5](#). Note that $x^2 + 2$ is irreducible over $\mathbb{Q}$, so no Jordan form exists over $\mathbb{Q}$ and Exercise (C1: A matrix is similar to its transpose) would not have decided this pair. (Here $\mathbf{B} \neq \mathbf{A}^\top$, so [Corollary 10.7.7](#) does not apply directly; the Smith forms settle it anyway.)

10.7.6.3 C. Going deeper

Exercise. C1: The same algorithm over the integers

The proof of [Theorem 10.7.3](#) used only one property of $F[x]$: division with remainder, where the “size” of a nonzero element is its degree. The integers $\mathbb{Z}$ have the same property, with $|a|$ in place of $\deg$.

- (a) State the version of [Theorem 10.7.3](#) for $\mathbf{M} \in M_n(\mathbb{Z})$ that the same proof gives, saying what replaces “monic”.
- (b) Compute the Smith normal form over $\mathbb{Z}$ of $\mathbf{M} = \begin{pmatrix} 2 & 4 \\ 6 & 4 \end{pmatrix}$, naming the operations.
- (c) Every finite abelian group G with n generators is $\mathbb{Z}^n$ modulo the column span of some $\mathbf{M} \in M_n(\mathbb{Z})$, the matrix of relations, exactly as V was $F[x]^n$ modulo the columns of $x\mathbf{I} - \mathbf{A}$. Using (b), identify the group with relation matrix $\mathbf{M}$ as a product of cyclic groups.

Hint: for (c), a diagonal relation matrix $\text{diag}(d_1, \dots, d_n)$ presents the product of the cyclic groups $\mathbb{Z}/d_i\mathbb{Z}$.

Solution.

- (a) Every $\mathbf{M} \in M_n(\mathbb{Z})$ can be brought by integer elementary operations — swap two rows or columns, multiply a row or column by ± 1 (the units of $\mathbb{Z}$), add an **integer** multiple of one row or column to another — to a diagonal matrix $\text{diag}(d_1, \dots, d_n)$ with every $d_i \geq 0$ and $d_i \mid d_{i+1}$, and this diagonal is unique. “Monic” is replaced by “nonnegative”, which is the normalization available because the units of $\mathbb{Z}$ are ± 1 . Uniqueness again follows from the determinantal divisors Δ_k , now nonnegative integer gcds of the $k \times k$ minors, using the Cauchy–Binet formula (Theorem 7.7.2, valid over any commutative ring by the remark after it) over the commutative ring $\mathbb{Z}$.
- (b) The least absolute value among the entries is 2, already in the corner. It divides 4 and 6. Subtract 2 times column 1 from column 2, and 3 times row 1 from row 2:

$$\begin{pmatrix} 2 & 4 \\ 6 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 0 \\ 6 & -8 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 0 \\ 0 & -8 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix},$$

the last step multiplying row 2 by -1 . Since $2 \mid 8$, the Smith normal form is $\text{diag}(2, 8)$. Check: $\Delta_1 = \gcd(2, 4, 6, 4) = 2$ and $\Delta_2 = |\det \mathbf{M}| = |8 - 24| = 16 = 2 \cdot 8$.

- (c) By (b) and the hint, the group is $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$, of order $16 = |\det \mathbf{M}|$. This is the structure theorem for finite abelian groups in miniature: the invariant divisors $d_1 \mid \dots \mid d_n$ of the relation matrix are the orders of the cyclic factors.

Exercise. C2: Uniqueness by hand in the smallest case

Let $\mathbf{M} \in M_2(F[x])$ be nonzero.

- (a) Prove directly, without using determinantal divisors, that $\Delta_1(\mathbf{M})$ — the monic gcd of the four entries of $\mathbf{M}$ — is unchanged by each of the six elementary operations of Definition 10.7.1.
- (b) Prove that $\det(\mathbf{UMV}) = c \det \mathbf{M}$ for a nonzero constant c whenever $\mathbf{U}, \mathbf{V}$ are invertible over $F[x]$.
- (c) Deduce that if $\text{diag}(d_1, d_2) \approx \text{diag}(e_1, e_2)$ with all four entries monic or zero, $d_1 \mid d_2$, and $e_1 \mid e_2$, then $d_1 = e_1$ and $d_2 = e_2$.

Solution.

- (a) Write $d = \Delta_1(\mathbf{M})$, the monic gcd of the entries, that is, the monic generator of the ideal $\langle m_{11}, m_{12}, m_{21}, m_{22} \rangle$ (Theorem 6.2.2). A swap permutes the entries, so the ideal is unchanged. Scaling a row by $c \in F \setminus \{0\}$ replaces two generators g by cg , and $\langle cg \rangle = \langle g \rangle$ since c is a unit, so the ideal is unchanged. Adding h times row 1 to row 2 replaces m_{2j} by $m_{2j} + hm_{1j}$, which lies in the old ideal; so the new ideal is contained in the old one, and the operation is reversed by adding $-h$ times row 1 to row 2, which gives the reverse containment. The same three arguments apply to columns. Hence d is unchanged.
- (b) By multiplicativity of the determinant over the commutative ring $F[x]$ (Theorem 7.5.4

and the remark on commutative rings), $\det(\mathbf{UMV}) = \det \mathbf{U} \cdot \det \mathbf{M} \cdot \det \mathbf{V}$, and $\det \mathbf{U}, \det \mathbf{V}$ are nonzero constants by [Proposition 10.7.1](#). Take $c = \det \mathbf{U} \det \mathbf{V}$.

- (c) By [Proposition 10.7.2](#) and (a), Δ_1 is an invariant of equivalence as far as elementary operations go; and for a general equivalence $\mathbf{M}' = \mathbf{UMV}$ the argument of (a) applies verbatim with the generators of the ideal replaced using $m'_{ij} = \sum_{k,l} u_{ik} m_{kl} v_{lj} \in \langle m_{kl} \rangle$, and symmetrically. For the two diagonal matrices, $\Delta_1(\text{diag}(d_1, d_2)) = \gcd(d_1, d_2) = d_1$ since $d_1 \mid d_2$ and d_1 is monic (or $d_1 = d_2 = 0$); likewise $\Delta_1(\text{diag}(e_1, e_2)) = e_1$. Hence $d_1 = e_1$. By (b), $d_1 d_2 = c e_1 e_2$ for some nonzero constant c ; if $d_1 = e_1 \neq 0$ then $d_2 = c e_2$, and two monic polynomials (or two zeros) differing by a constant factor are equal, so $d_2 = e_2$. If $d_1 = e_1 = 0$, then $d_2 = e_2 = 0$ by the divisibility chain.