Chapter 4
Operators

4.1 Ore Algebras and Ore Actions

Chapters 2 and 3 have been written to highlight the parallels between differential and recurrence equations. We have seen that most things (definitions, theorems, algorithms, etc.) laid out in one chapter have a natural counterpart in the other chapter. Our goal is now to develop a more general theory that includes both differential equations and recurrence equations as special cases by adopting the viewpoint of operators. In fact, we have already used differential operators and recurrence operators without formally introducing them: they were viewed as polynomials $p_0 + p_1 X + \cdots + p_r X^r$, where X played the role of derivation or shift, multiplied by coefficients p_i. In order to get the desired property that $(LM) \cdot f = L \cdot (M \cdot f)$, i.e., that the product of two operators acts on a function in the same way as the two factors act in succession, we were forced to give up commutativity of the multiplication. For example, if we write a function f very explicitly in the form $(t \mapsto f(t))$, then we have $x \cdot (D \cdot f) = (t \mapsto tf'(t))$ and $D \cdot (x \cdot f) = (t \mapsto tf'(t) + f(t))$, so the operators xD and Dx cannot be equal. Instead, we need that $Dx = xD + 1$. Similarly, we have $x \cdot (S \cdot f) = (t \mapsto tf(t+1))$ and $S \cdot (x \cdot f) = (t \mapsto (t+1)f(t+1))$, so the operators xS and Sx cannot be equal either. Instead, we need to have $Sx = (x+1)S$. These so-called *commutation rules* motivate the following definition.

Definition 4.1 Let R be a ring.

1. If $\sigma\colon R \to R$ is a ring endomorphism and $\delta\colon R \to R$ is such that $\delta(a+b) = \delta(a) + \delta(b)$ and $\delta(ab) = \delta(a)b + \sigma(a)\delta(b)$ for all $a, b \in R$, then δ is called a σ-*derivation*. The subset $\mathrm{Const}(R) = \mathrm{Const}_{\sigma,\delta}(R) = \{c \in R : \sigma(c) = c \wedge \delta(c) = 0\}$ of R is called the *constant ring* of R (with respect to σ and δ).
2. Suppose that a ring structure is defined on the set $R[X]$ of univariate polynomials in X with coefficients in R, and suppose that its addition agrees with the usual addition and its multiplication is such that $X^i X^j = X^{i+j}$ $(i, j \in \mathbb{N})$ and there

M. Kauers, *D-Finite Functions*, Algorithms and Computation in Mathematics 30,
https://doi.org/10.1007/978-3-031-34652-1_4

is an endomorphism $\sigma\colon R \to R$ and a σ-derivation $\delta\colon R \to R$ such that $Xa = \sigma(a)X + \delta(a)$ for all $a \in R$. Suppose further that the multiplication in $R[X]$ extends the multiplication of R in the sense that $R[X]$ is a left-R-module. Then $R[X]$ is called an *Ore algebra* over R. □

Since the multiplication in $R[X]$ need not be commutative, we should insist on the convention that polynomials have the form $p_0 + p_1X + \cdots + p_rX^r$, i.e., the coefficients p_i are placed on the left side of the terms X^i. Note that whenever a polynomial is given in some other form (e.g., with all coefficients on the left, or even with mixed terms such as pX^iqX^jr), a repeated application of the commutation rules always allow us to bring them into the standard form where all coefficients are on the left (see also Theorem 4.3 below). This form is unique because the powers of X are understood to be linearly independent over R.

Example 4.2

1. The ring $C[x][D]$ of differential operators is an Ore algebra. This ring is known as the first Weyl algebra. We have $\sigma = \mathrm{id}$ and $\delta = \frac{d}{dx}$. Also, the ring $C(x)[D]$ of differential operators is an Ore algebra for which we have $\sigma = \mathrm{id}$ and $\delta = \frac{d}{dx}$. Even more generally, if R is any differential ring, then $R[D]$ is an Ore algebra with $\sigma = \mathrm{id}$ and δ being the derivation of R.
   In $C[x][D]$ we have, for example,

$$\begin{aligned}(aD+b)(cD+d) &= (aD+b)cD + (aD+b)d\\ &= a(cD+c')D + bcD + a(dD+d') + bd\\ &= acD^2 + (ac' + bc + ad)D + (ad' + bd),\end{aligned}$$

   where c', d' refer to the derivatives of c, d.
2. The Euler derivation is defined as $\theta := x\frac{d}{dx}$. The ring $C[x][\theta]$ of Euler-differential operators is an Ore algebra with $\sigma = \mathrm{id}$ and $\delta = x\frac{d}{dx}$.
3. The ring $C[x][S]$ of recurrence operators is an Ore algebra. In this case, σ is the function that maps $p(x) \in C[x]$ to $p(x+1)$, and δ is identically zero. Again, we may take $C(x)$ instead of $C[x]$, and, more generally, if R is any difference ring, then $R[S]$ is an Ore algebra with σ being the endomorphism of R and δ being identically zero.
   For example, in $C[x][S]$ we have

$$\begin{aligned}(aS+b)(cS+d) &= aScS + aSd + bcS + bd\\ &= a\sigma(c)S^2 + (a\sigma(d) + bc)S + bd.\end{aligned}$$

4. Let $q \in C$ be fixed and define the q-shift S_q by $(S_qf)(x) = f(qx)$. The set $C[x][S_q]$ of all q-shift-recurrence operators $p_0 + p_1S_q + \cdots + p_rS_q^r$ forms an Ore algebra with $\sigma\colon C[x] \to C[x]$ as the function that maps $p(x)$ to $p(qx)$ and δ being identically zero. Since $C[x]$ together with σ is a difference ring,

this example is just another special case of an Ore algebra $R[S]$ with R being a difference ring.

5. Another special case is the Ore algebra $C[x][M_q]$ of Mahler operators. Here, for a fixed $q \in \mathbb{N}$ the operator M_q is defined through $(M_q f)(x) = f(x^q)$. The corresponding ring of linear operators $p_0 + p_1 M_q + \cdots + p_r M_q^r$ forms an Ore algebra with $\sigma : C[x] \to C[x]$ being the function that maps $p(x)$ to $p(x^q)$ and with δ being the zero function.
6. As an alternative to recurrence equations, we can consider difference equations, which are expressed in terms of the forward difference operator Δ defined via $(\Delta f)(x) = f(x+1) - f(x)$. The ring $C[x][\Delta]$ of all operators of the form $p_0 + p_1\Delta + \cdots + p_r\Delta^r$ is an Ore algebra with the shift as σ (as in the previous case) and the function that maps $p(x) \in C[x]$ to $p(x+1) - p(x)$ as δ. It can be checked (Exercise 1) that this δ is indeed a σ-derivation.
7. Let $C = \mathbb{Q}(\alpha)$ be an algebraic number field and let σ be an element of the Galois group of C over $\mathbb{Q}$. Then there is an Ore algebra $C[X]$ with the commutation rule $Xa = \sigma(a)X$ for all $a \in C$.
   More specifically, take $C = \mathbb{C}$ and let σ be the conjugation map, i.e., $\sigma(a+b\mathrm{i}) = \overline{a+b\mathrm{i}} = a - b\mathrm{i}$ for any $a, b \in \mathbb{R}$. Then there is an Ore algebra $\mathbb{C}[X]$ with the commutation rule $Xa = \bar{a}X$ for all $a \in \mathbb{C}$.
8. Let C be a field of characteristic $p > 0$ and let $\sigma : C \to C$ be the Frobenius endomorphism defined by $\sigma(a) = a^p$ for all $a \in C$. Then there is an Ore algebra $C[X]$ with the commutation rule $Xa = \sigma(a)X$ for all $a \in C$.
9. It is not required that the ground ring R is commutative or free of zero divisors. For example, let $R = C^{2\times 2}$, $A = \left(\begin{smallmatrix} 2 & 3 \\ 1 & 4 \end{smallmatrix}\right)$, and define $\delta : R \to R$ by $\delta(M) = MA - AM$. Then δ is a derivation on R and there is an Ore algebra $R[X]$ whose commutation rule is $Xa = aX + \delta(a)$. □

Are all of the Ore algebras mentioned in the example above really well-defined? On the one hand, it is easy to see that a commutation rule $Xa = \sigma(a)X + \delta(a)$ can only work if σ is an endomorphism and δ is a σ-derivation. To see this, just observe that

$$Xab = (Xa)b = (\sigma(a)X + \delta(a))b = \sigma(a)\sigma(b)X + (\delta(a)b + \sigma(a)\delta(b)),$$
$$Xab = X(ab) = \sigma(ab)X + \delta(ab),$$

so $\sigma(ab) = \sigma(a)\sigma(b)$ and $\delta(ab) = \delta(a)b + \sigma(a)\delta(b)$ for all $a, b \in R$. So the restrictions on σ and δ are necessary. On the other hand, it could be questioned whether the conditions imposed on σ and δ are also sufficient, i.e., whether every choice of σ and δ indeed gives rise to an Ore algebra. The following theorem asserts that this is the case.

Theorem 4.3 *Let R be a ring, $\sigma : R \to R$ an endomorphism, and $\delta : R \to R$ a σ-derivation. Then there exists exactly one Ore algebra $R[X]$ with $Xa = \sigma(a)X + \delta(a)$ for all $a \in R$.* □

Proof We already know that $R[X]$ together with the usual addition forms an abelian group. We have to show that there is a unique way to define a multiplication on $R[X]$ that is compatible with this addition and satisfies the required commutation rule. To be compatible means that it should be associative and that it should satisfy the two distributive laws.

Distributivity enforces

$$\left(\sum_{i=0}^{n} a_i X^i\right)\left(\sum_{j=0}^{m} b_j X^j\right) = \sum_{i=0}^{n}\sum_{j=0}^{m}(a_i X^i)(b_j X^j)$$

for any choice of $a_0, \ldots, a_n, b_0, \ldots, b_m \in R$. For each of the terms $(a_i X^i)(b_j X^j)$, associativity enforces $(a_i X^i)(b_j X^j) = a_i(X^i b_j)X^j$. The desired commutation rule enforces $X^i b_j = X^{i-1}(Xb_j) = X^{i-1}\sigma(b_j)X + X^{i-1}\delta(b_j)$, and, by induction on i, that there is at most one choice of $c_0, \ldots, c_i \in R$ such that $X^i b_j = c_0 + c_1 X + \cdots + c_i X^i$. Since $X^i X^j = X^{i+j}$ for all $i, j \in \mathbb{N}$ and the multiplication of coefficients must agree with the multiplication of the ground ring, it follows that we must have $(a_i X^i)(b_j X^j) = (a_i c_0)X^j + (a_i c_1)X^{1+j} + \cdots + (a_i c_i)X^{i+j}$. It is therefore shown that there is *at most* one Ore algebra for a given pair (σ, δ).

For the existence, first define inductively $\gamma(u, k, n)$ for every $u \in R$, every $n \in \mathbb{N}$, and every $k \in \mathbb{Z}$ by $\gamma(u, 0, 0) = u$, $\gamma(u, k, 0) = 0$ for $k \neq 0$, $\gamma(u, k, n) = 0$ for $k < 0$, and $\gamma(u, k, n+1) = \sigma(\gamma(u, k-1, n)) + \delta(\gamma(u, k, n))$. Note that this definition implies $\gamma(u, k, n) = 0$ for $k > n$. We define $X^n u := \sum_k \gamma(u, k, n)X^k$ for every $n \in \mathbb{N}$ and set

$$\left(\sum_{i=0}^{n} a_i X^i\right)\left(\sum_{j=0}^{m} b_j X^j\right) := \sum_{i=0}^{n}\sum_{j=0}^{m} a_i(X^i b_j)X^j.$$

The distributive laws and the commutation rule $Xu = \sigma(u)X + \delta(u)$ $(u \in R)$ are then satisfied by construction, and it remains to check associativity. Because of distributivity, it suffices to check that $((aX^i)(bX^j))(cX^k) = (aX^i)((bX^j)(cX^k))$ for all $a, b, c \in R$ and $i, j, k \in \mathbb{N}$. In fact, it even suffices to check $(XbX^k)c = X(bX^k c)$ for all $b, c \in R$ and all $k \in \mathbb{N}$, because the case $i > 1$ can be treated by induction, and multiplying with an element of R from the left or with a power of X from the right is harmless. For all $b, c \in R$ and $k \in \mathbb{N}$ we have

$$\begin{aligned}(XbX^k)c &= \sigma(b)X^{k+1}c + \delta(b)X^k c\\ &= \sum_j \big(\sigma(b)\gamma(c, j, k+1) + \delta(b)\gamma(c, j, k)\big)X^j\\ &= \sum_j \big(\sigma(b)\sigma(\gamma(c, j-1, k)) + \sigma(b)\delta(\gamma(c, j, k)) + \delta(b)\gamma(c, j, k)\big)X^j,\end{aligned}$$

$$
\begin{aligned}
X(bX^k c) &= \sum_j Xb\gamma(c, j, k)X^j \\
&= \sum_j \sigma(b\gamma(c, j, k))X^{j+1} + \delta(b\gamma(c, j, k))X^j \\
&= \sum_j \big(\sigma(b)\sigma(\gamma(c, j-1, k)) + \delta(b)\gamma(c, j, k) + \sigma(b)\delta(\gamma(c, j, k))\big)X^j,
\end{aligned}
$$

and since both quantities agree, the proof is complete. ■

Implicit in the above proof is the following multiplication algorithm for elements of Ore algebras.

Algorithm 4.4
Input: *Two elements* $P = p_0 + p_1 X + \cdots + p_r X^r$ *and* Q *of an Ore algebra* $R[X]$
Output: *The product* $PQ \in R[X]$

1 *Set* $R = 0$.
2 *for* $i = 0, \ldots, r$ *do*
3 $R = R + p_i Q$ *(here* p_i *is multiplied to the left of each coefficient of* Q*).*
4 $Q = \sigma(Q)X + \delta(Q)$ *(here the understanding is that* σ, δ *are applied to the coefficients of* Q*).*
5 *Return* R.

Some authors use notations like $R\langle X\rangle$ to emphasize the non-commutativity of an Ore algebra. Others write $R[X; \sigma, \delta]$ in order to include the two functions governing the commutation rules into the notation. We will stick to the common notation $R[X]$ for univariate polynomials, but instead of X we will often use the symbol ∂ (not to be confused with δ) for denoting the indeterminate. With this notation, the elements of $R[\partial]$ look more like operators. Note however that the formal construction does not require that the elements of $R[\partial]$ are operators, we just use them primarily for that purpose.

The commutation rule strongly restricts the non-commutativity of an Ore algebra, to the effect that Ore algebras are more closely related to commutative polynomial rings than to more general non-commutative rings.

Definition 4.5 Let $R[\partial]$ be an Ore algebra.

1. The *order* of a nonzero element $L \in R[\partial]$ is defined as the largest $r \in \mathbb{N}$ such that the coefficient $[\partial^r]L$ of ∂^r in L is nonzero. It is denoted by $\operatorname{ord}(L)$. We also define $\operatorname{ord}(0) = -\infty$.
2. For an element $L \in R[\partial] \setminus \{0\}$, the coefficient $\operatorname{lc}(L) := [\partial^{\operatorname{ord}(L)}]L$ is called the *leading coefficient,* and $\operatorname{lt}(L) = \partial^{\operatorname{ord}(L)}$ is called the *leading term*. $L \in R[\partial]$ is called *monic* if $\operatorname{lc}(L) = 1$.
3. A nonzero element L of a left ideal I of $R[\partial]$ is called *minimal* if $\operatorname{ord}(M) \geq \operatorname{ord}(L)$ for all nonzero elements M of I. □

Proposition 4.6 *Let* $R[\partial]$ *be an Ore algebra.*

1. *We have* $\operatorname{ord}(ML) = \operatorname{ord}(M) + \operatorname{ord}(L)$ *for all* $M, L \in R[\partial]$ *if and only if* σ *is injective and* R *is an integral domain.*
2. *Suppose that* σ *is injective and* R *is an integral domain. Then every left ideal* $I \neq \{0\}$ *of* $R[\partial]$ *has a unique monic minimal element if and only if* R *is in fact a field.* □

Proof

1. "⇒": For $p \in \ker\sigma \setminus \{0\}$, we have $\partial p = \sigma(p)\partial + \delta(p) = \delta(p)$, so $\operatorname{ord}(\partial p) < \operatorname{ord}(\partial) + \operatorname{ord}(p)$ in contradiction to the assumption. Also, for any $u, v \in R \setminus \{0\}$ with $uv = 0$, we have $\operatorname{ord}(uv) < \operatorname{ord}(u) + \operatorname{ord}(v)$ in contradiction to the assumption.
   "⇐": For any $L = \ell_n\partial^n + \cdots$, $M = m_k\partial^k + \cdots$ in $R[\partial]$ we have $LM = \ell_n\sigma^n(m_k)\partial^{n+k} + \cdots$, and the assumptions guarantee that $\ell_n\sigma^n(m_k)$ is nonzero if ℓ_n and m_k are nonzero.
2. "⇒": If R is not a field, then it has some non-invertible element $p \in R$. The left ideal $\langle p\rangle$ generated by p in $R[\partial]$ contains an element of order 0 (namely p), but no monic element of order 0 (i.e., it does not contain 1). To see this, observe that every element of $\langle p\rangle$ has the form $Lp = (\ell_n\partial^n + \cdots + \ell_0)p = \ell_n\sigma^n(p)\partial^n + \cdots$, and by the assumptions on R and σ, the coefficient $\ell_n\sigma^n(p)$ is not zero when ℓ_n is not zero. So in order for Lp to have order 0, we must have $n = 0$, but then $Lp = \ell_0 p$ cannot be 1 because p was assumed not to be invertible in R.
   "⇐": If L is any element of I for which $\operatorname{ord}(L)$ is minimal, then $\operatorname{lc}(L)^{-1}L$ is a monic minimal element of I, so monic minimal elements always exist. If L_1, L_2 are two monic minimal elements of I, then $\operatorname{ord}(L_1) = \operatorname{ord}(L_2)$ and $\operatorname{lc}(L_1) = \operatorname{lc}(L_2) = 1$ implies that $\operatorname{ord}(L_1 - L_2) < \operatorname{ord}(L_1) = \operatorname{ord}(L_2)$, which by minimality of L_1, L_2 implies that $L_1 - L_2 = 0$, i.e., $L_1 = L_2$. ■

From a formal perspective, the elements of an Ore algebra $R[\partial]$ become operators as soon as we have a left-$R[\partial]$-module F on which they can act. In view of the applications we have in mind, we call such modules *function spaces.* Recall that to be a left-$R[\partial]$-module means that there is an addition $+\colon F \times F \to F$ which turns F into an abelian group and a multiplication $\cdot\colon R[\partial] \times F \to F$ which satisfies the rules

$$\begin{aligned} 1 \cdot f &= f, \\ (L + M) \cdot f &= (L \cdot f) + (M \cdot f), \\ L \cdot (f + g) &= (L \cdot f) + (L \cdot g), \\ (LM) \cdot f &= L \cdot (M \cdot f), \end{aligned}$$

for all $L, M \in R[\partial]$ and all $f, g \in F$. For example, the ring $C[[x]]$ of formal power series is a left-$C[x][D]$-module if we define multiplication via

$$(p_0 + p_1 D + \cdots + p_r D^r) \cdot f = p_0 f + p_1 f' + \cdots + p_r f^{(r)}.$$

Definition 4.7 Let $R[\partial]$ be an Ore algebra and F be a left-$R[\partial]$-module.

1. For $f \in F$, we call $\operatorname{ann}(f) = \{ L \in R[\partial] : L \cdot f = 0 \}$ the *annihilator* of f (in $R[\partial]$).
2. $f \in F$ is *D-finite* (with respect to the action of $R[\partial]$ on F) if $\operatorname{ann}(f) \neq \{0\}$.
3. For $L \in R[\partial]$, we call $V(L) = \{ f \in F : L \cdot f = 0 \}$ the *solution space* of L (in F). □

Example 4.8

1. Taking $C[x][D]$ as $R[\partial]$ and $C[[x]]$ as F, we have

$$D - 1 \in \operatorname{ann}(\exp(x)) \quad \text{and} \quad \exp(x) \in V(D-1),$$

because $(D-1) \cdot \exp(x) = 0$. Note that we also have $5\exp(x) \in V(D-1)$ and $xD - x \in \operatorname{ann}(\exp(x))$.
2. Taking $C[x][S]$ as $R[\partial]$ and $C^{\mathbb{N}}$ as F, we have

$$S - 2 \in \operatorname{ann}((2^n)_{n=0}^{\infty}) \quad \text{and} \quad (2^n)_{n=0}^{\infty} \in V(S-2),$$

because $(S-2) \cdot (2^n)_{n=0}^{\infty} = (0)_{n=0}^{\infty}$. We also have $(c\,2^n)_{n=0}^{\infty} \in V(S-2)$ for any $c \in C$ and $xS - 2x \in \operatorname{ann}((2^n)_{n=0}^{\infty})$, but, for example, $S - x \notin \operatorname{ann}((2^n)_{n=0}^{\infty})$, because $(S-x) \cdot (2^n)_{n=0}^{\infty} = (2^{n+1} - n2^n)_{n=0}^{\infty} = ((2-n)2^n)_{n=0}^{\infty} \neq (0)_{n=0}^{\infty}$.
3. The definition of D-finiteness is compatible with the definitions we gave earlier for the shift and differential case, and it covers other cases as well. For example, if we let the Ore algebra $C(x)[M_2]$ of Mahler operators act on $C((x))$, we find that $(xM_2 - 1) \cdot \frac{1}{x} = x\frac{1}{x^2} - \frac{1}{x} = 0$, so $\frac{1}{x}$ is D-finite. Other examples such as $\exp(x)$ are not D-finite in this setting (Exercise 13), and $\sum_{n=0}^{\infty} x^{2^n}$ is D-finite with respect to $C(x)[M_2]$ (an annihilating operator is $M_2^2 - 2M_2 + 1$) but it is not D-finite with respect to $C(x)[D]$ (Exercise 3 in Sect. 3.1).
4. If $R[\partial]$ is an Ore algebra, then $R[\partial]$ is naturally a left-module over itself. Also R can be viewed as a left-$R[\partial]$-module. If we set $\partial \cdot 1 := u$ for some element $u \in R$ of our choice, then $\partial \cdot r = (\partial r) \cdot 1 = (\sigma(r)\partial + \delta(r)) = \sigma(r)u + \delta(r)$ fixes the action. As a concrete example, think of the natural action of $C[x][D]$ on $C[x]$. Here we have $D \cdot 1 = 0$, which implies that $D \cdot r = \delta(r)$ for all $r \in C[x]$. On the other hand, for the natural action of $C[x][S]$ on $C[x]$, we have $S \cdot 1 = 1$, which together with $\delta = 0$ implies $S \cdot r = \sigma(r)$ for all $r \in C[x]$. □

Theorem 4.9 *Let $R[\partial]$ be an Ore algebra and F be a left-$R[\partial]$-module. Let $f \in F$ and $L \in R[\partial]$.*

1. *$\operatorname{ann}(f)$ is a left ideal of $R[\partial]$.*
2. *$V(L)$ is a $\operatorname{Const}(R)$-submodule of F.* □

Proof Both parts of the proof depend on the observation that $L \cdot 0 = 0$ for all $L \in R[\partial]$, which follows from the calculation $L \cdot 0 = L \cdot (0+0) = (L \cdot 0) + (L \cdot 0)$.

1. Clearly $\operatorname{ann}(f)$ is not empty because $0 \cdot f = 0$, so $0 \in \operatorname{ann}(f)$. Let $L_1, L_2 \in \operatorname{ann}(f)$ and $M_1, M_2 \in R[\partial]$. Then $(M_1L_1+M_2L_2)\cdot f = ((M_1L_1)\cdot f)+((M_2L_2)\cdot f) = (M_1 \cdot (L_1 \cdot f)) + (M_2 \cdot (L_2 \cdot f)) = (M_1 \cdot 0) + (M_2 \cdot 0) = 0 + 0 = 0$, so $M_1L_1 + M_2L_2 \in \operatorname{ann}(f)$.
2. Clearly $V(L)$ is not empty because $L \cdot 0 = 0$, so $0 \in V(L)$. Let $f, g \in V(L)$ and $\alpha, \beta \in \operatorname{Const}(R)$. Write $L = p_0 + p_1\partial + \cdots + p_r\partial^r$ with $p_0, \ldots, p_r \in R$, so that $L \cdot f = L \cdot g = 0$. Then $L \cdot ((\alpha \cdot f) + (\beta \cdot g)) = ((L\alpha) \cdot f) + ((L\beta) \cdot g) = (\alpha \cdot (L \cdot f)) + (\beta \cdot (L \cdot g)) = (\alpha \cdot 0) + (\beta \cdot 0) = 0 + 0 = 0$, so $(\alpha \cdot f) + (\beta \cdot g) \in V(L)$. ∎

The second part of the theorem looks more familiar if we apply it to the typical situation where $R = C(x)$ or $R = C[x]$ and an Ore algebra $R[\partial]$ where ∂ commutes with all elements of C but not with x. If $\partial = D$ or $\partial = S$, we have $\operatorname{Const}(R) = C$ and the statement reduces to the fact that $V(L)$ is a C-vector space. In general, the solution space $V(L)$ is not closed under multiplication by x or under application of ∂.

For any element $f \in F$, we can consider the left-$R[\partial]$-module generated by f in F. This is the set of all elements of F which can be written in the form $L \cdot f$ for some $L \in R[\partial]$. Let us denote this submodule of F by $R[\partial] \cdot f$. It is often the case that we want to know something about this submodule, e.g., whether it contains an element with a certain desired property. However, it is not particularly handy to view it as a submodule of F, as the elements of F are typically inherently infinite objects such as formal power series with which we cannot easily do computations. We can fix this by considering the module homomorphism $\phi\colon R[\partial] \to F$ defined by $\phi(L) = L \cdot f$. By the homomorphism theorem, we have $R[\partial]/\operatorname{ann}(f) \cong R[\partial] \cdot f$, and therefore, computations in $R[\partial] \cdot f$ are equivalent to computations in $R[\partial]/\operatorname{ann}(f)$. In most of what follows, we will be doing computations in an Ore algebra $R[\partial]$ or a quotient $R[\partial]/I$ for some left-ideal I of $R[\partial]$ rather than computations with explicit "functions".

Example 4.10 For $f = 1/(1 - \sqrt{x}) \in C[[x]]$ we have

$$(3f + x(x^2 - 1)f')' = -\tfrac{1}{2}(x + 1)f + \tfrac{1}{2}(x^2 - 4x + 5)f'.$$

The series f is annihilated by the operator $L = 2x(x - 1)D^2 + (5x - 1)D + 1 \in C(x)[D]$, and in $C(x)[D]/\langle L\rangle$ we have

$$D \cdot [3 + x(x^2 + 1)D] = [-\tfrac{1}{2}(x + 1) + \tfrac{1}{2}(x^2 - 4x + 5)D].$$

Note that the element $[1] \in C(x)[D]/\langle L\rangle$ plays the role of the function f which is annihilated by L. □

Left-$R[\partial]$-modules generalize the notion of D-modules introduced at the end of Sect. 3.2 and are sometimes also called *∂-modules* for short. Whenever we say $R[\partial]$-module, we always mean a left module.

Proposition 4.6 suggests to restrict the attention to Ore algebras $K[\partial]$ where K is a field. While every Ore algebra $R[\partial]$ where R is an integral domain can be uniquely

extended to an Ore algebra $\mathrm{Quot}(R)[\partial]$ (Exercise 5), not every $R[\partial]$-module admits a natural extension to a $\mathrm{Quot}(R)[\partial]$-module, so while the theory and algorithms are simpler for Ore algebras over fields, restricting the attention to these algebras is "with loss of generality".

Example 4.11

1. $C[[x]]$ is a $C[x][D]$-module but not a $C(x)[D]$-module, because not every element of $C[[x]]$ can be multiplied with any element of $C(x)$. However, $C((x))$ is a natural extension of $C[[x]]$ which is a $C(x)[D]$-module, and whenever $L \in C[x][D]$ is an annihilating operator of a series $f \in C[[x]] \subseteq C((x))$, then so is every element of the ideal generated by L in $C(x)[D]$. There is therefore no harm in working with $C(x)[D]$ instead of $C[x][D]$.
2. $C^{\mathbb{N}}$ is a $C[x][S]$-module but not a $C(x)[S]$-module, because we cannot meaningfully multiply a sequence with a rational function that has a pole at a nonnegative integer. It can happen that an operator L annihilates a sequence f, but not every $C(x)$-multiple of L does. For example, $L = (x-5)$ annihilates the sequence $f : \mathbb{N} \to C$ with $f(5) = 1$ and $f(n) = 0$ for $n \neq 5$, but $\frac{1}{x-5}L = 1 \in C[x][S]$ does not. So unlike in the first example, if $L \in C[x][S]$ annihilates a sequence $f \in C^{\mathbb{N}}$, the ideal generated by L in $C(x)[S]$ may contain operators that do not annihilate f, and they may even belong to $C[x][S]$. □

When it is appropriate to work with a field, the arguments given in earlier chapters for D-finite closure properties can be easily lifted to the general setting. They reduce to linear algebra over K.

Theorem 4.12 *Let $K[\partial]$ be an Ore algebra over a field K, and let F be a $K[\partial]$-module.*

1. *$f \in F$ is D-finite if and only if the dimension of $K[\partial] \cdot f$ as a K-vector space is finite. If f is D-finite and L is a minimal element of* $\mathrm{ann}(f)$*, then* $\dim_K(K[\partial] \cdot f) = \mathrm{ord}(L)$.
2. *If $f \in F$ is D-finite with respect to $K[\partial]$ and annihilated by an operator of order at most r, then the same is true for $M \cdot f \in F$, for every $M \in K[\partial]$.*
3. *If $f, g \in F$ are D-finite with respect to $K[\partial]$ and annihilated by operators of orders at most r and s, respectively, then $f+g$ is D-finite as well and annihilated by an operator of order at most $r+s$.* □

Proof

1. "⇒": Let L be the monic minimal element of $\mathrm{ann}(f)$, and let $r = \mathrm{ord}(L)$. We show that $K[\partial] \cdot f$ is generated by $f, (\partial \cdot f), \ldots, (\partial^{r-1} \cdot f)$. It suffices to show that every subspace generated by $f, (\partial \cdot f), \ldots, (\partial^k \cdot f)$, for some $k \geq r$, is generated by $f, (\partial \cdot f), \ldots, (\partial^{r-1} \cdot f)$, and this is easily seen by induction on k using that $\partial^k \cdot f = (\partial^k \cdot f) - \partial^{k-r} \cdot (L \cdot f) = (\partial^k - \partial^{k-r}L) \cdot f$ is contained in the subspace generated by $f, (\partial \cdot f), \ldots, (\partial^{k-1} \cdot f)$, for any $k \geq r$.
"⇐": If the dimension of $K[\partial] \cdot f$ is r, then any $r+1$ elements of $K[\partial] \cdot f$ must be linearly dependent over K. In particular, there will be $\ell_0, \ldots, \ell_r \in K$, not all

zero, such that $\ell_0 f + \cdots + \ell_r(\partial^r \cdot f) = 0$, i.e., $\operatorname{ann}(f)$ contains the nonzero element $L = \ell_0 + \cdots + \ell_r \partial^r$ and hence f is D-finite.
For the claim about the dimension, note that the argument given above for "$\Rightarrow$" already implies that $\dim_K(K[\partial] \cdot f) \le r$. If the dimension were smaller, then $f, \ldots, (\partial^{r-1} \cdot f)$ would be linearly dependent over K, and the linear dependence would give rise to a nonzero annihilating operator of order less than r, in contradiction to the minimality of L.

2. Let L be a minimal element of $\operatorname{ann}(f)$, and let $r = \operatorname{ord}(L)$. By part 1, the submodule $K[\partial] \cdot f$ of F is a K-vector space of dimension r. Since it is closed under ∂, it contains $M \cdot f$ and all its derivatives. Hence, the elements $M \cdot f, \ldots, \partial^r \cdot (M \cdot f)$ of $K[\partial] \cdot f$ are linearly dependent over K, i.e., there are $p_0, \ldots, p_r \in K$, not all zero, such that $(p_0 + \cdots + p_r \partial^r) \cdot (M \cdot f) = 0$, as claimed.
3. $f + g$ is an element of the submodule $(K[\partial] \cdot f) + (K[\partial] \cdot g) \subseteq F$. By part 1, this submodule is a K-vector space of dimension at most $r + s$, hence $(f + g), \ldots, \partial^{s+r} \cdot (f + s)$ must be linearly dependent, and the dependence gives rise to an annihilating operator for $f + g$. ∎

We have seen in the previous chapters that D-finiteness is also preserved under multiplication (cf. Theorems 2.30 and 3.25). In order to lift this property to the general realm of Ore algebras, we must consider function spaces F that have a multiplication, and the multiplication must be compatible with the module structure. By a multiplication we mean a K-bilinear function

$$m\colon F \times F \to F, \qquad m(f, g) = fg,$$

i.e., a function which is additive in both arguments and satisfies $pm(f, g) = m(pf, g) = m(f, pg)$ for all $f, g \in F$ and all $p \in K$. It need not be commutative or associative, nor is it necessary to have a neutral element in F. But we do want to have a product rule that relates m to the action of ∂. More precisely, we will assume that there are $\alpha, \beta, \gamma \in K$ such that for all $f, g \in F$ we have

$$\partial \cdot m(f, g) = \alpha m(f, g) + \beta m(\partial \cdot f, g) + \beta m(f, \partial \cdot g) + \gamma m(\partial \cdot f, \partial \cdot g).$$

If f, g are D-finite, say $L \cdot f = M \cdot g = 0$ for some nonzero operators $L, M \in K[\partial]$, then every $\partial^n \cdot m(f, g)$ ($n \in \mathbb{N}$) belongs to the K-vector space generated by $m(\partial^i \cdot f, \partial^j \cdot g)$ for $i = 0, \ldots, \operatorname{ord}(L) - 1$ and $j = 0, \ldots, \operatorname{ord}(M) - 1$ in F. As this subspace has dimension at most $\operatorname{ord}(L)\operatorname{ord}(M)$, we find that $m(f, g), \ldots, \partial^{\operatorname{ord}(L)\operatorname{ord}(M)} \cdot m(f, g)$ are linearly dependent over K and thus $m(f, g)$ is D-finite and annihilated by an operator of order at most $\operatorname{ord}(L)\operatorname{ord}(M)$. The argument is always the same: a system of linear equations with more variables than equations must have a nontrivial solution.

It is worth noting that in the case of addition, all we need to know for computing an annihilating operator for $f + g$ are annihilating operators L and M of f and g. In fact, we do not need any f, g, F to begin with, but can start right away from an Ore

algebra $K[\partial]$ and two nonzero elements L and M. We can then choose the module $F = (K[\partial]/\langle L\rangle) \times (K[\partial]/\langle M\rangle)$ with the action $P \cdot ([v],[w]) = ([Pv],[Pw])$. Then for $f = ([1],[0])$ and $g = ([0],[1])$ we have $L \cdot f = M \cdot g = 0$, and since $\dim_K(F) = \operatorname{ord}(L) + \operatorname{ord}(M)$, we find that $f + g = ([1],[1])$ is annihilated by an operator of order (at most) $\operatorname{ord}(L) + \operatorname{ord}(M)$. This operator will be such that for every $K[\partial]$-module F and any elements $f, g \in F$ with $L \cdot f = M \cdot g = 0$, it annihilates $f + g$.

Also for multiplication, we do not need to know much about the module F or its multiplication function m. All that enters into the argument are the coefficients $\alpha, \beta, \gamma \in K$ that connect ∂ to m. Given two operators $L, M \in K[\partial]$, we can use the vector space tensor product $(K[\partial]/\langle L\rangle) \otimes_K (K[\partial]/\langle M\rangle)$ as the function space F. Recall from linear algebra that the tensor product of two K-vector spaces V, W consists of all finite K-linear combinations of the formal quantities $v \otimes w$ with $v \in V$ and $w \in W$, which satisfy the laws $v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$, $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$ and $p(v \otimes w) = (pv) \otimes w = v \otimes (pw)$ for all $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$ and $p \in K$. With the help of $\alpha, \beta, \gamma \in K$, we can turn the K-vector space $F = (K[\partial]/\langle L\rangle) \otimes_K (K[\partial]/\langle M\rangle)$ into a $K[\partial]$-module, by defining $\partial \cdot ([v] \otimes [w]) = \alpha([v] \otimes [w]) + \beta([\partial v] \otimes [w]) + \beta([v] \otimes [\partial w]) + \gamma([\partial v] \otimes [\partial w])$. The coefficients α, β, γ are not uniquely determined by $K[\partial]$ (Exercise 19), but also not completely arbitrary (Exercise 20).

Definition 4.13 Let $K[\partial]$ be an Ore algebra and let $L, M \in K[\partial]$ be nonzero.

1. Let $\alpha, \beta, \gamma \in K$ be such that the definition

$$\partial \cdot ([v] \otimes [w]) := \alpha([v] \otimes [w]) + \beta([\partial v] \otimes [w]) + \beta([v] \otimes [\partial w]) + \gamma([\partial v] \otimes [\partial w])$$

   for $v, w \in K[\partial]$ turns $F = (K[\partial]/\langle L\rangle) \otimes_K (K[\partial]/\langle M\rangle)$ into a $K[\partial]$-module. Then the unique monic minimal element of $\operatorname{ann}([1] \otimes [1])$ is called the *symmetric product* of L and M with respect to α, β, γ. It is denoted by $L \otimes M$.
2. With $\alpha, \beta, \gamma \in K$ as above we define $L^{\otimes 1} = L$ and $L^{\otimes(n+1)} = L \otimes L^{\otimes n}$ for $n \geq 1$, and call $L^{\otimes n}$ the nth *symmetric power* of L with respect to α, β, γ.
3. Now let $F = (K[\partial]/\langle L\rangle) \times (K[\partial]/\langle M\rangle)$ and turn F into a $K[\partial]$-module by setting $\partial \cdot ([v],[w]) := ([\partial v],[\partial w])$ for all $v, w \in K[\partial]$. Let $s = ([1],[1])$. Then the unique monic minimal element of $\operatorname{ann}(s)$ is called the *symmetric sum* of L and M. It is denoted by $L \oplus M$. □

It is not hard to see that the symmetric sum and the symmetric product are commutative, i.e., we have $L \otimes M = M \otimes L$ and $L \oplus M = M \oplus L$ for all $M, L \in K[\partial]$. Furthermore, we have $1 \otimes M = 1$ and $1 \oplus M = M$ for all $M \in K[\partial]$, in agreement with the fact that we must have $0f = 0$ and $0 + f = f$ for any element f of any $K[\partial]$-module F. Finally, $\otimes$ and $\oplus$ are associative (Exercise 22), and we have the distributive law $L \otimes (M_1 \oplus M_2) = (L \otimes M_1) \oplus (L \otimes M_2)$ for $L, M_1, M_2 \in K[\partial]$ (Exercise 23). The Ore algebra $K[\partial]$ together with the operations $\oplus$ and $\otimes$ is a commutative semi-ring. It is not a ring because it lacks a notion of subtraction.

Example 4.14 Symmetric sums and products can be computed by linear algebra. For example, consider the Ore algebra $K[\partial]$ with $K = C(x)$, and with $\sigma, \delta\colon K \to K$ defined by $\sigma(p(x)) = p(x^2)$ and $\delta(p(x)) = 5p(x^2) - 5p(x)$ for all $p(x) \in K$. Let $L = \partial + x^2$ and $M = \partial^2 - x$.

1. We have

$$L \oplus M = \partial^3 + \frac{x^{12}+x^{11}-4x^{10}-4x^9-9x^8-9x^7+16x^6+16x^5+26x^4+26x^3-25x^2-25x+5}{x^4+x^3-4x^2-4x+1}\partial^2 - x^2\partial + \frac{-x^{13}-x^{12}+4x^{11}+4x^{10}+9x^9+9x^8-16x^7-21x^6-26x^5-x^4+25x^3}{x^4+x^3-4x^2-4x+1}.$$

This operator can be found as follows. First, apply successive powers of ∂ to the element $([1], [1])$ of the product space $K[\partial]/\langle L\rangle \times K[\partial]/\langle M\rangle$. Note that whenever an operator of order ≥ 1 appears in a first component, we can subtract from it a suitable left-multiple of L and replace it by a representative of order < 1. For the second component, any representative of order ≥ 2 can be replaced by a representative of order < 2 by adding a suitable left-multiple of M. In other words, every element of $K[\partial]/\langle L\rangle \times K[\partial]/\langle M\rangle$ is a K-linear combination of $([1], [0])$, $([0], [1])$, and $([0], [\partial])$. In particular,

$$\begin{aligned}
([1], [1]) &= ([1], [1]),\\
\partial \cdot ([1], [1]) &= ([-x^2], [\partial]),\\
\partial^2 \cdot ([1], [1]) &= ([x^6 - 5x^4 + 5x^2], [x]),\\
\partial^3 \cdot ([1], [1]) &= ([-x^{14} + 5x^{12} + 5x^{10} - 25x^8 - 10x^6 + 50x^4 - 25x^2],\\
&\qquad [x^2\partial + 5x^2 - 5x]).
\end{aligned}$$

Now we make an ansatz for an operator $P = p_0 + p_1\partial + p_2\partial^2 + p_3\partial^3$ and set up a linear system to enforce $P \cdot ([1], [1]) = ([0], [0])$. Coefficient comparison leads to

$$\begin{pmatrix} 1 & -x^2 & x^6-5x^4+5x^2 & -x^{14}+5x^{12}+5x^{10}-25x^8-10x^6+50x^4-25x^2 \\ 1 & 0 & x & 5x^2-5x \\ 0 & 1 & 0 & x^2 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} = 0$$

whose solution space is generated by the coefficient vector of the operator announced above.

2. For the choice $\alpha = 20$, $\beta = 5$, $\gamma = 1$, we have

$$L \otimes M = \partial^2 + 10(x^4 - 4)\partial - (x^7 - 25x^6 - 5x^5 + 75x^4 - 5x^3 + 125x^2 + 25x - 400).$$

In order to find this operator, we find a linear relation among the elements $\partial^n \cdot (1 \otimes 1)$ $(n = 0, 1, 2)$ of the tensor product space $K[\partial]/\langle L\rangle \otimes_K K[\partial]/\langle M\rangle$. We have

$$\begin{aligned}
[1] \otimes [1] &= 1([1] \otimes [1]) + 0([1] \otimes [\partial]),\\
\partial \cdot ([1] \otimes [1]) &= 20([1] \otimes [1]) + 5([-x^2] \otimes [1])\\
&\quad + 5([1] \otimes [\partial]) + ([-x^2] \otimes [\partial])\\
&= (-5x^2 + 20)([1] \otimes [1]) + (5 - x^2)([1] \otimes [\partial]),\\
\partial^2 \cdot ([1] \otimes [1]) &= (x^7 + 25x^6 - 5x^5 - 125x^4 - 5x^3\\
&\quad - 75x^2 + 25x + 400)([1] \otimes [1])\\
&\quad - (10x^6 - 50x^4 - 40x^2 + 200)([1] \otimes [\partial]).
\end{aligned}$$

Now we make an ansatz for an operator $P = p_0 + p_1\partial + p_2\partial^2$ and set up a linear system to enforce $P \cdot ([1] \otimes [1]) = ([0] \otimes [0])$. Coefficient comparison leads to a linear system for p_0, p_1, p_2 whose solution space is generated by the coefficient vector of the operator announced above. □

Typically, we will not have a natural interest in the modules $(K[\partial]/\langle L\rangle) \times (K[\partial]/\langle M\rangle)$ or $(K[\partial]/\langle L\rangle) \otimes_K (K[\partial]/\langle M\rangle)$, but we want to reason about elements of some other $K[\partial]$-modules F. For example, there might be two specific D-finite elements f, g of F for which we know annihilating operators $L, M \in K[\partial]$, and we might want to know an annihilating operator of their sum $f + g$. The point is that although $(K[\partial]/\langle L\rangle) \times (K[\partial]/\langle M\rangle)$ may be different from F, we can be sure that the symmetric sum $L \oplus M$ will be an annihilating operator of $f + g$. The reason is that when L and M are annihilating operators of f and g, respectively, then we can define a module homomorphism

$$\phi\colon (K[\partial]/\langle L\rangle) \times (K[\partial]/\langle M\rangle) \to F, \quad \phi([U], [V]) = (U \cdot f) + (V \cdot g),$$

and so if $P \in K[\partial]$ is an annihilating operator for $([1], [1])$, i.e., $P \cdot ([1], [1]) = 0$, then also $\phi(P \cdot ([1], [1])) = P \cdot \phi([1], [1]) = P \cdot (f + g) = 0$. The reasoning for multiplication is analogous, except that in this case F must also be a K-algebra, and α, β, γ must be chosen in such a way that they are compatible with the multiplication of F.

The symmetric sum and the symmetric product can thus be used to work in arbitrary $K[\partial]$-modules F, and they are oblivious to the particular F that we have in mind. A price we have to pay for this generality is that in general we cannot preserve minimality: even if L and M are the unique monic minimal operators annihilating f and g, respectively, the operators $L \oplus M$ and $L \otimes M$ are in general not minimal annihilating operators of $f + g$ and fg, respectively. As a counterexample, consider

the case $g = -f$ and $M = L$. In this case, we have $L \oplus L = L$ (Exercise 21) while $f + g = 0$ is also annihilated by $1 \in K[\partial]$.

For the multiplication case, take for example $f = 1 + \exp(x)$, $g = 1 - \exp(x) \in C((x))$ and $L = M = D^2 - D$. Then we have $L \otimes M = D^3 - 3D^2 + 2D$ but $fg = 1 - \exp(2x)$ is also annihilated by $D^2 - 2D$.

Exercises

1. Let $\sigma, \delta \colon C[x] \to C[x]$ be defined by $\sigma(p(x)) = p(x+1)$ for $p \in C[x]$ and $\delta(p(x)) = p(x+1) - p(x)$. Show that δ is a σ-derivation.

2. Write the elements $(x+\partial^2)(1-2\partial+x\partial^2)$ and $1-\partial(x+3x^2)+(x+1)\partial^2(x-1)$ of $K[\partial]$ in the standard form $p_0 + p_1\partial + p_2\partial^2 + \cdots$, given that

a. $\sigma = \mathrm{id}$ and $\delta = \frac{d}{dx}$,

b. $\sigma(p(x)) = p(x+1)$ and $\delta = 0$,

c. $\sigma(p(x)) = p(x^2)$ and $\delta(p(x)) = 5p(x^2) - 5p(x)$.

3⋆. Show that $D^n x^k = \sum_{i \geq 0} \binom{n}{i} k^{\underline{i}} x^{k-i} D^{n-i}$ for all $n, k \in \mathbb{N}$.

4⋆. Let R be an integral domain, $\sigma \colon R \to R$ be an endomorphism, and $\delta \colon R \to R$ be a σ-derivation. For $m \in \mathbb{N}$ and $p \in R$, we use the notation $\sigma^{\overline{m}}(p) = p\sigma(p)\cdots\sigma^{m-1}(p)$. Show that we have

$$\delta(\sigma^{\overline{m}}(p)) = \delta\big(p + \sigma(p) + \cdots + \sigma^{m-1}(p)\big)\sigma^{\overline{m-1}}(\sigma(p))$$

for all $m \in \mathbb{N}$ and all $p \in R$. This formula generalizes the formula $D(a^n) = na^{n-1}D(a)$ from Exercise 9 in Sect. 3.2.

5. Let R be an integral domain, $\sigma \colon R \to R$ be an injective endomorphism, and $\delta \colon R \to R$ be a σ-derivation. Show that for $K = \mathrm{Quot}(R)$, there exists exactly one endomorphism $\bar{\sigma} \colon K \to K$ and exactly one $\bar{\sigma}$-derivation $\bar{\delta} \colon K \to K$ with $\bar{\sigma}|_R = \sigma$ and $\bar{\delta}|_R = \delta$.

6. Show that in a Laurent Ore polynomial ring $R[\partial, \partial^{-1}]$, we must have $\partial^{-1}\sigma(p) = p\partial^{-1} - \partial^{-1}\delta(p)\partial^{-1}$ for all $p \in R$.

7. Let $L \in K[\partial]$ and $m, y \in K$. Prove or disprove:

a. $L \cdot y = m \Rightarrow (L - m) \cdot y = 0$.

b. $(L - m) \cdot y = 0 \Rightarrow L \cdot y = m$.

8. Let $\sigma \colon K \to K$ be an endomorphism and $\delta \colon K \to K$ be a σ-derivation. Show:

a. If $\sigma \neq \mathrm{id}$, then there exists an element $u \in K$ such that $\delta(q) = u(\sigma(q) - q)$ for all $q \in K$.

b. If $\delta \neq 0$, then there exists an element $u \in K$ such that $\sigma(q) = u\delta(q) + q$ for all $q \in K$.

9. Prove or disprove: If $\sigma\colon R \to R$ is an endomorphism and $\delta\colon R \to R$ is a σ-derivation, then $\sigma \circ \delta = \delta \circ \sigma$.

10. Let R be an integral domain, and $\sigma \neq \mathrm{id}$ or $\delta \neq 0$. Suppose that σ is injective. Let $Z(R[\partial]) = \{\, p \in R[\partial] \mid \forall\, q \in R[\partial] : pq = qp \,\}$ be the centralizer of $R[\partial]$, and consider $\mathrm{Const}(R) \subseteq R \subseteq R[\partial]$. Prove or disprove:

a. $\mathrm{Const}(R) \subseteq Z(R[\partial])$

b. $Z(R[\partial]) \subseteq \mathrm{Const}(R)$

11. Determine the unique monic minimal annihilating operator of $\frac{x+1}{x-1} \in C(x)$ **a.** in $C(x)[D]$; **b.** in $C(x)[S]$; **c.** in $C(x)[M_2]$.

12. Is the commutative polynomial ring $C(x)[Y]$ an Ore algebra? Can we view $F = C((x))$ as a $C(x)[Y]$-module with the action $Y^i \cdot f = f^i$, so that a series is D-finite with respect to $C(x)[Y]$ if and only if it is algebraic?

13⋆. Show that $\exp(x)$ is not D-finite with respect to $C(x)[M_2]$.

14. Show that $f(x) = \sum_{n=0}^{\infty} q^{n^2} x^n$ with q not a root of unity is D-finite with respect to the q-shift operator but not with respect to the usual derivation.

Hint: You may use without proof that for any pairwise distinct $\phi_1, \ldots, \phi_r \in C$ the sequences $(\phi_i^n)_{n=0}^{\infty}$ are linearly independent over $C(n)$.

15. Prove or disprove: For all $L, M \in C(x)[D]$ there exists $\tilde{M} \in C(x)[D]$ such that $LM = \tilde{M}L$.

16. Show that a sequence is D-finite with respect to $C[x][S]$ if and only if it is D-finite with respect to $C[x][\Delta]$.

17⋆. Let $R[\partial]$ be an Ore algebra over a commutative ring R. Show that for all $L, M \in R[\partial] \setminus \{0\}$ we have $\mathrm{ord}(LM) = \mathrm{ord}(L) + \mathrm{ord}(M)$ if and only if R is an integral domain and σ is injective.

18⋆. For $F = C((x))$, the Hadamard product $m\colon F \times F \to F$, $m(f, g) = f \odot g$ is a C-bilinear function. Show that there are no $\alpha, \beta, \gamma \in C$ such that for all $f, g \in C((x))$ we have $m(f, g)' = \alpha m(f, g) + \beta m(f', g) + \beta m(f, g') + \gamma(f', g')$.

19. Let F be a $K[\partial]$-module and $m\colon F \times F \to F$ a bilinear map such that there are $\alpha, \beta, \gamma \in K$ with $\partial \cdot m(f, g) = \alpha m(f, g) + \beta m(\partial \cdot f, g) + \beta m(f, \partial \cdot g) + \gamma m(\partial \cdot f, \partial \cdot g)$ for all $f, g \in F$. Let $q \in K \setminus \{0\}$ and define $\tilde{m}\colon F \times F \to F$ by $\tilde{m}(f, g) := qm(f, g)$. Determine $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma} \in K$ such that $\partial \cdot \tilde{m}(f, g) = \alpha \tilde{m}(f, g) + \beta \tilde{m}(\partial \cdot f, g) + \beta \tilde{m}(f, \partial \cdot g) + \gamma \tilde{m}(\partial \cdot f, \partial \cdot g)$ for all $f, g \in F$.

20⋆⋆. Show that if $\alpha, \beta, \gamma \in K$ are such that the symmetric product with respect to α, β, γ is well-defined for an Ore algebra $K[\partial]$, then $(p - \sigma(p))\alpha + \delta(p)(\beta - 1) = (p - \sigma(p))\beta + \delta(p) = 0$ for all $p \in K$.

21. Show that for all $M, L \in K[\partial]$ we have $L \oplus M \in \langle L \rangle \cap \langle M \rangle$. Conclude that $L \oplus L = L$ for all monic $L \in K[\partial]$.

22. Show that the operations $\otimes, \oplus\colon K[\partial] \times K[\partial] \to K[\partial]$ are associative.

23. Let $\alpha, \beta, \gamma \in K$ be such that the symmetric product with respect to α, β, γ is defined for an Ore algebra $K[\partial]$. Show that for all $M, L_1, L_2 \in K[\partial]$ we have $M \otimes (L_1 \oplus L_2) = (M \otimes L_1) \oplus (M \otimes L_2)$.

24**. (Stavros Garoufalidis and Christoph Koutschan) Let $C(q)$ be a field of rational functions over C, consider the Ore algebra $C(q)(x)[Q]$ with $\sigma(p(x)) = p(qx)$ for $p \in C(q)(x)$ and $\delta = 0$, and let $F = C(q)^{\mathbb{N}}$ be the set of sequences in $C(q)$. The set F becomes a $C(q)(x)[Q]$-module by setting $x \cdot (a_n(q))_{n=0}^{\infty} = (q^n a_n(q))_{n=0}^{\infty}$ and $Q \cdot (a_n(q))_{n=0}^{\infty} = (a_{n+1}(q))_{n=0}^{\infty}$. Let $\omega \in C$ be a root of unity. Show that if $(a_n(q))_{n=0}^{\infty}$ is D-finite with respect to the action of $C(q)(x)[Q]$, then so is $(a_n(\omega q))_{n=0}^{\infty}$.

Hint: Show that for any integer $k \geq 2$, there is an annihilating operator of $(a_n(q))_{n=0}^{\infty}$ with polynomial coefficients in which all exponents of x are divisible by k.

25. Let the Ore algebra $C(x)[M_2]$ act on $C((x))$ via $M_2 \cdot f(x) = f(x^2)$ for $f \in C((x))$. Suppose that $f, g \in C((x))$ are such that $(M_2^2 + xM_2 + x^2) \cdot f = (M_2^2 - xM_2 + x^2) \cdot g = 0$. Compute annihilating operators for $f + g$ and fg.

26. Let $K[\partial]$ be an Ore algebra and F be a $K[\partial]$-module. Let $L \in K[\partial] \setminus \{0\}$ and let $f, g \in F$ be such that $L \cdot f = g$. Prove or disprove:

a. If f is D-finite, then so is g.

b. If g is D-finite, then so is f.

27. (Clemens Raab) Let F be a $K[\partial]$-module and $m\colon F \times F \to F$ be a bilinear map with $m(f, g) = m(g, f)$ for all $f, g \in F$. Let $L, M \in K[\partial]$ be such that $L \cdot f = 0 \Rightarrow M \cdot m(f, f) = 0$ for all $f \in F$. Show that then we even have $M \cdot m(f_1, f_2) = 0$ for any two $f_1, f_2 \in F$ with $L \cdot f_1 = L \cdot f_2 = 0$.

References

General expositions on the theory of noncommutative rings can be found in [186, 303, 316]. As remarked in the text, the rings we consider here are, in a way, only slightly noncommutative, which makes it possible to handle them without first going through a general course on noncommutative rings. Ore algebras were introduced and first studied by Ore [344]. He already drew his motivation from differential and recurrence operators. Bronstein and Petkovšek introduced Ore algebras into computer algebra in their tutorial paper [115].

It has been remarked that the orders of $L \oplus M$ and $L \otimes M$ are in general larger than necessary. At the same time, it can be shown that the orders do not overshoot if the module F is sufficiently large, in the following sense: if $L, M \in C(x)[D]$ are such that $\dim_C V(L) = \operatorname{ord}(L)$ and $\dim_C V(M) = \operatorname{ord}(M)$, then $\dim_C V(L \oplus M) = \operatorname{ord}(L \oplus M)$ and $V(L \oplus M) = V(L) + V(M)$. Moreover, if F is even a differential ring, then $\dim_C V(L \otimes M) = \operatorname{ord}(L \otimes M)$ and $V(L \otimes M)$ is the C-vector space generated by $\{ fg : f \in V(L), g \in V(M)\}$ in F. Proofs can be found in a paper of

Singer [406] or, using more abstract constructions, in the book of van der Put and Singer [441].

The multiplication algorithm stated in this section is straightforward. More sophisticated algorithms are known, at least for the differential case. Benoit, Bostan, and van der Hoeven [54] showed that the product of two elements of $C[x][D]$ whose degrees in x and D are d and r, respectively, can be computed with $\mathrm{O}^{\sim}(\min(d,r)^{\omega-2}dr)$ operations in C, where ω is the exponent of matrix multiplication (cf. Sect. 1.4). Bostan, Chyzak, and Le Roux [84] showed some sort of converse: the product of two $n \times n$ matrices with coefficients C can be computed with a number of operations in C that does not exceed the number of operations in C needed to compute the a certain fixed number of products of two elements of $C[x][D]$ whose degrees in both x and D are bounded by n.

4.2 Common Right Divisors and Left Multiples

In this section, we consider an arbitrary Ore algebra $K[\partial]$ over a field K. We assume throughout that σ is injective. We have already observed that despite being non-commutative, the Ore algebra $K[\partial]$ has some similarities with the commutative polynomial ring $C[x]$. For example, the order in $K[\partial]$ plays the role of the degree in $C[x]$. Thanks to the degree function, $C[x]$ is a Euclidean domain, and we will now see that $K[\partial]$ is a right-Euclidean domain.

Theorem 4.15 *For every $U, V \in K[\partial]$ with $v \neq 0$ there exists a unique pair $(Q, R) \in K[\partial]^2$ such that $U = QV + R$ and* $\operatorname{ord}(R) < \operatorname{ord}(V)$. □

Proof There clearly exist Q, R with $U = QV + R$, for example $Q = 0, R = U$ is a valid choice. Among all pairs (Q, R) with $U = QV + R$, select one for which $\operatorname{ord}(R)$ is minimal. We show that $\operatorname{ord}(R) < \operatorname{ord}(V)$. If we had $\operatorname{ord}(R) \geq \operatorname{ord}(V)$, there is a $c \in K$ such that $\operatorname{ord}(R - c\partial^{\operatorname{ord}(R)-\operatorname{ord}(V)}V) < \operatorname{ord}(R)$, and for $R' = R - c\partial^{\operatorname{ord}(R)-\operatorname{ord}(V)}V$ and $Q' = Q + c\partial^{\operatorname{ord}(R)-\operatorname{ord}(V)}$ we have $U = Q'V + R'$, in contradiction to the minimality assumption on R.

We have thus shown that a pair (Q, R) with $U = QV + R$ and $\operatorname{ord}(R) < \operatorname{ord}(V)$ always exists. For the uniqueness, suppose there is another pair (Q', R') with $U = Q'V + R'$ and $\operatorname{ord}(R') < \operatorname{ord}(V)$. Then $(Q - Q')V = R' - R$. The left hand side has order $\operatorname{ord}(Q - Q') + \operatorname{ord}(V)$, while the order of the right hand side is strictly less than $\operatorname{ord}(V)$. Therefore $\operatorname{ord}(Q - Q') < 0$, which means $Q = Q'$. But then $0 = (Q - Q')V = R' - R$ also implies $R' = R$. ■

Definition 4.16 Let $U, V \in K[\partial]$, $V \neq 0$, and let $Q, R \in K[\partial]$ be as in Theorem 4.15. Then $\operatorname{rquo}(U, V) := Q$ is called the *right quotient* and $\operatorname{rrem}(U, V) := R$ is called the *right remainder* of U with respect to V. If $R = 0$, we say that V is a *right factor* or *right divisor* of U and that U is a *left multiple* of V. □

Given two elements $U, V \in K[\partial]$, we can compute the right quotient and the right remainder in very much the same way as in the commutative case. The proof of Theorem 4.15 translates into the following algorithm.

Algorithm 4.17
Input: $U, V \in K[\partial]$ *with* $V \neq 0$.
Output: $\operatorname{rquo}(U, V)$ *and* $\operatorname{rrem}(U, V)$.

Let $Q = 0$ *and* $R = U$.
while $\operatorname{ord}(R) > \operatorname{ord}(V)$, *do*
  $c = \frac{\operatorname{lc}(R)}{\sigma^{\operatorname{ord}(R)-\operatorname{ord}(V)}(\operatorname{lc}(V))}$
  $R = R - c\partial^{\operatorname{ord}(R)-\operatorname{ord}(V)}V$
  $Q = Q + c\partial^{\operatorname{ord}(R)-\operatorname{ord}(V)}$
Return (Q, R).

Example 4.18

1. For $U = (3x+5)D^3 - (2x+1)D^2 + (2x-3)D + (3x+1)$ and $V = (x+2)D - (3x+5) \in C(x)[D]$ the algorithm yields

$$\operatorname{rquo}(U, V) = \frac{3x+5}{x+2}D^2 + \frac{7x^2+19x+13}{(x+2)^2}D + \frac{23x^3+108x^2+177x+100}{(x+2)^3},$$

$$\operatorname{rrem}(U, V) = \frac{72x^4+479x^3+1212x^2+1374x+586}{(x+2)^3}.$$

   It can be seen that $\operatorname{ord}(\operatorname{rrem}(U, V)) = 0 < 1 = \operatorname{ord}(V)$, and it can be checked that $U = \operatorname{rquo}(U, V)V + \operatorname{rrem}(U, V)$.
2. For $U = (3x+5)S^3 - (2x+1)S^2 + (2x-3)S + (3x+1)$ and $V = (x+2)S - (3x+5) \in C(x)[S]$ the algorithm yields

$$\operatorname{rquo}(U, V) = \frac{3x+5}{x+4}S^2 + \frac{7x^2+39x+51}{(x+3)(x+4)}S + \frac{23x^3+184x^2+468x+372}{(x+2)(x+3)(x+4)},$$

$$\operatorname{rrem}(U, V) = \frac{72x^4+695x^3+2411x^2+3554x+1884}{(x+2)(x+3)(x+4)}.$$

   Again it can be seen that $\operatorname{ord}(\operatorname{rrem}(U, V)) = 0 < 1 = \operatorname{ord}(V)$, and it can be checked that $U = \operatorname{rquo}(U, V)V + \operatorname{rrem}(U, V)$. Note that although U, V have the same coefficients as before, the results are not the same. The coefficients of $\operatorname{rquo}(U, V)$, $\operatorname{rrem}(U, V)$ depend on the arithmetic of $K[\partial]$, which is governed by σ and δ. □

Definition 4.19 Let $U, V \in K[\partial]$, not both zero.

1. If $G \in K[\partial]$ is a right divisor of both U and V, it is called a *common right divisor* of U and V.

2. A common right divisor G of U and V is called a *greatest common right divisor* if it is monic and a right divisor of any other common right divisor.
3. If $M \in K[\partial]$ is a left multiple of both U and V, it is called a *common left multiple* of U and V.
4. A common left multiple M of U and V is called a *least common left multiple* if it is monic and a right divisor of any other common left multiple. □

It is easy to show that for any pair $(U, V) \in K[\partial]^2 \setminus \{(0,0)\}$ there is at most one greatest common right divisor and at most one least common left multiple (Exercise 2). We denote these by $\operatorname{gcrd}(U, V)$ and $\operatorname{lclm}(U, V)$, respectively. We further define $\operatorname{gcrd}(0, 0) = 0$ and $\operatorname{lclm}(0, 0) = 0$.

Like in the commutative case, the existence of a greatest common right divisor follows from the correctness of the Euclidean algorithm, which happens to apply literally in the same way to Ore algebras. Also the extended Euclidean algorithm, which in addition to $\operatorname{gcrd}(U, V)$ computes $S, T \in K[\partial]$ such that $\operatorname{gcrd}(U, V) = SU + TV$, works for arbitrary Ore algebras $K[\partial]$.

Algorithm 4.20 *(Extended Euclidean Algorithm)*
Input: $U, V \in K[\partial]$, *not both zero.*
Output: $\operatorname{gcrd}(U, V)$ *and* $S, T \in K[\partial]$ *such that* $\operatorname{gcrd}(U, V) = SU + TV$.

Let $(G, S, T, G', S', T') = (U, 1, 0, V, 0, 1)$.
while $G' \neq 0$ *do*
  $Q = \operatorname{rquo}(G, G')$
  $(G, S, T, G', S', T') = (G', S', T', G - QG', S - QS', T - QT')$
Return $(\operatorname{lc}(G)^{-1}G, \operatorname{lc}(G)^{-1}S, \operatorname{lc}(G)^{-1}T)$.

Theorem 4.21 *Algorithm 4.20 is correct and terminates. In particular:*

1. *Any two elements* $U, V \in K[\partial]$ *with* $(U, V) \neq (0, 0)$ *have a greatest common right divisor* $\operatorname{gcrd}(U, V)$.
2. *For any* $U, V \in K[\partial]$ *with* $(U, V) \neq (0, 0)$ *there exist* $S, T \in K[\partial]$ *and* $\operatorname{gcrd}(U, V) = SU + TV$.
3. *If* $\operatorname{ord}(U) \geq \operatorname{ord}(V) \geq 0$ *and* $\operatorname{lc}(V)U \neq \operatorname{lc}(U)V$, *then for the pair* $(S, T) \in K[\partial]^2$ *computed by Algorithm 4.20 we have* $\operatorname{ord}(S) < \operatorname{ord}(V) - \operatorname{ord}(G)$ *and* $\operatorname{ord}(T) < \operatorname{ord}(U) - \operatorname{ord}(G)$.
4. *For any* $U, V \in K[\partial]$ *with* $(U, V) \neq (0, 0)$ *there exists at most one pair* $(S, T) \in K[\partial]^2$ *with* $\operatorname{gcrd}(U, V) = SU + TV$ *and* $\operatorname{ord}(S) < \operatorname{ord}(V) - \operatorname{ord}(G)$ *and* $\operatorname{ord}(T) < \operatorname{ord}(U) - \operatorname{ord}(G)$. □

Proof Termination is clear because $G - QG' = \operatorname{rrem}(G, G')$ implies that the order of G' decreases in every iteration, and since it is a natural number, it cannot decrease infinitely often. For the correctness, let $C(A, B) \subseteq K[\partial]$ denote the set of common right divisors of $A, B \in K[\partial]$. We then have $C(A, B) = C(B, A - QB)$ for every $Q \in K[\partial]$, because if D is a common right divisor of A and B, say $A = \tilde{A}D$ and $B = \tilde{B}D$ for some $\tilde{A}, \tilde{B} \in K[\partial]$, then $A - QB = (\tilde{Q} - Q\tilde{B})D$, so D is a common

right divisor of B and $A - QB$. This shows $C(A, B) \subseteq C(B, A - QB)$, and the other inclusion follows by symmetry (replace Q by $-Q$).

We have thus shown that $C(U, V) = C(G, G')$ at the end of every iteration. Upon termination, we have $G' = 0$, and since $C(G, 0)$ contains exactly the right divisors of G, the monic element $\mathrm{lc}(G)^{-1}G$ must be the greatest common right divisor of U and V. For the claim about S and T, observe first that we have $G = SU + TV$ and $G' = S'U + T'V$ at the beginning, after every iteration of the while loop, and therefore right after the while loop.

This proves the correctness of the algorithm. Parts 1 and 2 of the theorem follow immediately. For part 3, consider first the case when $\mathrm{ord}(U) > \mathrm{ord}(V)$ and let S, T be as computed by Algorithm 4.20. Define $(G_0, S_0, T_0) = (U, 1, 0)$, and let (G_k, S_k, T_k) be the values of G, S, T at the end of the kth iteration ($k = 1, 2, \ldots$). If Q_k denotes the value of Q in the kth iteration, we have $Q_k = \mathrm{rquo}(G_{k-1}, G_k)$ for all $k \geq 1$, which implies $\mathrm{ord}(Q_k) = \mathrm{ord}(G_{k-1}) - \mathrm{ord}(G_k)$ for all $k \geq 1$. Therefore $\mathrm{ord}(G_k) = \mathrm{ord}(G_1) - \sum_{i=1}^{k} \mathrm{ord}(Q_i)$ for all $k \geq 1$. By the definition of G_k and the assumption $\mathrm{ord}(U) > \mathrm{ord}(V)$, we have $\mathrm{ord}(Q_k) > 0$ for all $k \geq 1$. Therefore, from $S_{k+1} = S_{k-1} - Q_k S_k = S_{k-1} - Q_k(S_{k-2} + Q_{k-1}S_{k-1})$ it follows that $\mathrm{ord}(S_{k+1}) = \mathrm{ord}(Q_k) + \mathrm{ord}(S_k)$ for all $k \geq 2$. Taking also into account that $S_2 = S_0 - Q_1 S_1 = 1$, we obtain $\mathrm{ord}(S_k) = \sum_{i=2}^{k-1} \mathrm{ord}(Q_i)$.

Suppose now that the algorithm terminates after the kth iteration, so that $G_k = G = \mathrm{gcrd}(U, V)$, $S_k = S$, $T_k = T$. Because of the assumption $\mathrm{ord}(U) > \mathrm{ord}(V) \geq 0$, we must have $k \geq 2$. Therefore, $\mathrm{ord}(G) = \mathrm{ord}(G_1) - \sum_{i=1}^{k} \mathrm{ord}(Q_i) < \mathrm{ord}(V) - \sum_{i=2}^{k-1} \mathrm{ord}(Q_i) = \mathrm{ord}(V) - \mathrm{ord}(S)$, so $\mathrm{ord}(S) < \mathrm{ord}(V) - \mathrm{ord}(G)$. Moreover, since $\mathrm{ord}(G) \leq \mathrm{ord}(V) < \mathrm{ord}(U)$ and we must have $S \neq 0$ when $k \geq 2$, the equation $G = SU + TV$ implies $\mathrm{ord}(S) + \mathrm{ord}(U) = \mathrm{ord}(T) + \mathrm{ord}(V)$, from which we obtain $\mathrm{ord}(T) = \mathrm{ord}(U) + \mathrm{ord}(S) - \mathrm{ord}(V) < \mathrm{ord}(U) - \mathrm{ord}(G)$. This completes the proof of the order bounds for S and T in the case $\mathrm{ord}(U) > \mathrm{ord}(V)$.

For the case $\mathrm{ord}(U) = \mathrm{ord}(V)$, we have $G_0 = U$, $G_1 = V$, $G_2 = V - \frac{\mathrm{lc}(V)}{\mathrm{lc}(U)}U$. The assumption on U and V ensures $G_2 \neq 0$ and $\mathrm{ord}(G_2) < \mathrm{ord}(V)$. We can therefore apply the previous argument with V and G_2 in place of U and V and obtain $S', T' \in K[\partial]$ with $G = S'V + T'G_2 = S'V + T'(V - \frac{\mathrm{lc}(V)}{\mathrm{lc}(U)}U) = (S' + T')V - T'\frac{\mathrm{lc}(U)}{\mathrm{lc}(V)}U$ and $\mathrm{ord}(S') < \mathrm{ord}(V) - \mathrm{ord}(G)$ and $\mathrm{ord}(T') < \mathrm{ord}(G_2) - \mathrm{ord}(G)$. Algorithm 4.20 applied to U and V will therefore give $S = T'\frac{\mathrm{lc}(U)}{\mathrm{lc}(V)}$ and $T = S' + T'$, and for these we have $\mathrm{ord}(S) = \mathrm{ord}(T') < \mathrm{ord}(G_2) - \mathrm{ord}(G) \leq \mathrm{ord}(V) - \mathrm{ord}(G)$ and $\mathrm{ord}(T) \leq \max(\mathrm{ord}(S'), \mathrm{ord}(T')) < \mathrm{ord}(V) - \mathrm{ord}(G) = \mathrm{ord}(U) - \mathrm{ord}(G)$.

The proof of part 4 is Exercise 6. ∎

Example 4.22

1. In $C(x)[D]$, consider the operators

$$U = (x-1)D^5 + 5D^4,$$

$$V = (x-1)D^3 + (6-3x)D^2 + (27x^2 + 9x - 42)D + (117 - 54x).$$

For the sequence of successive remainders, we have $G_0 = U$, $G_1 = V$ and then

$$\begin{aligned} G_2 &= \operatorname{rrem}(G_0, G_1) = -162(x^2-1)D^2 - 81(9x^3+12x^2-11x-18)D \\ &\qquad - 81(18x^2-21x-41), \\ G_3 &= \operatorname{rrem}(G_1, G_2) = \frac{27}{4}(3x+4)(x+2)(x-1)D - \frac{27(3x+4)(2x^2-x-7)}{4(x+1)}, \\ G_4 &= \operatorname{rrem}(G_2, G_3) = 0. \end{aligned}$$

It follows that

$$\operatorname{gcrd}(U, V) = \operatorname{lc}(G_3)^{-1} G_3 = D - \frac{2x^2-x-7}{(x+2)(x+1)(x-1)}.$$

2. In $C(x)[S]$, consider the operators

$$\begin{aligned} U &= S^7 + 5S^6 + 9S^5 + 5S^4 - 5S^3 - 9S^2 - 5S - 1, \\ V &= (x+5)S^5 + 6S^4 - 6S^3 - 2(x-1)S^2 - (x+19)S + 2(x+6). \end{aligned}$$

For the sequence of successive remainders, we have $G_0 = U$, $G_1 = V$ and then

$$\begin{aligned} G_2 &= \operatorname{rrem}(G_0, G_1) \\ &= +\frac{(x+4)(7x+27)}{(x+6)(x+7)}S^4 + \frac{2(3x^2+37x+102)}{(x+6)(x+7)}S^3 + \frac{2(x+5)(6x+1)}{(x+6)(x+7)}S^2 \\ &\quad - \frac{2(3x^2-13x-148)}{(x+6)(x+7)}S - \frac{19x+103}{x+7}, \\ G_3 &= \operatorname{rrem}(G_1, G_2) \\ &= -\frac{16(x+3)(x+7)(3x+11)}{(7x+27)(7x+34)}S^3 + \frac{16(x+7)(x^2-12x-58)}{(7x+27)(7x+34)}S^2 \\ &\quad + \frac{16(x+7)(3x^2+40x+103)}{(7x+27)(7x+34)}S - \frac{16(x+2)(x+6)(x+7)}{(7x+27)(7x+34)}, \\ G_4 &= \operatorname{rrem}(G_2, G_3) \\ &= \frac{2(x+2)(2x+7)(7x+27)(7x+34)}{(x+6)(x+7)(3x+11)(3x+14)}S^2 \\ &\quad + \frac{20(x+4)(7x+27)(7x+34)}{(x+6)(x+7)(3x+11)(3x+14)}S - \frac{2(2x+9)(7x+27)(7x+34)}{(x+7)(3x+11)(3x+14)}, \\ G_5 &= \operatorname{rrem}(G_3, G_4) = 0. \end{aligned}$$

It follows that

$$\operatorname{gcrd}(U, V) = \operatorname{lc}(g_4)^{-1} g_4 = S^2 + \frac{10(x+4)}{(x+2)(2x+7)} S - \frac{(2x+9)(x+6)}{(2x+7)(x+2)}.$$

□

The example illustrates a phenomenon that also appears in the commutative case: the intermediate successive remainders have much larger coefficients than the final result. Since we need them only up to nonzero K-multiples, it is a good idea to introduce an additional instruction $(G, S, T) = (\operatorname{lt}(G)^{-1} G, \operatorname{lt}(G)^{-1} S, \operatorname{lt}(G)^{-1} T)$ at the end of the loop body in order to clear useless common K-factors that blow up the coefficients.

The example also illustrates a phenomenon that does not appear in the commutative case: the coefficients of the greatest common right divisor can be larger than those of the input. In the case of integers or univariate commutative polynomial rings over a field, there is a lemma by Gauss which says that this cannot happen, but as the example above shows, there is no natural counterpart of this lemma for $K[\partial]$.

In order to get a bound on the degree of the coefficients in the greatest common right divisor, we translate the question into linear algebra. It suffices to consider $U, V \in K[\partial]$ which are both nonzero and such that $\operatorname{lc}(V)U \neq \operatorname{lc}(U)V$ (otherwise the greatest common right divisor is obvious). Under these assumptions, we know from Theorem 4.21 that $\operatorname{gcrd}(U, V)$ can be written as $SU + TV$ for certain $S, T \in K[\partial]$ with $\operatorname{ord}(S) < \operatorname{ord}(V)$ and $\operatorname{ord}(T) < \operatorname{ord}(U)$. Since we know $r_V := \operatorname{ord}(V)$ and $r_U := \operatorname{ord}(U)$ when we know U and V, we can make an ansatz $S = s_0 + s_1\partial + \cdots + s_{r_V-1}\partial^{r_V-1}$ and $T = t_0 + t_1\partial + \cdots + t_{r_U-1}\partial^{r_U-1}$ with undetermined coefficients $s_0, \ldots, s_{r_V-1}, t_0, \ldots, t_{r_U-1} \in K$. If we write $\operatorname{gcrd}(U, V) = g_0 + g_1\partial + \cdots + g_{r_U+r_V-1}\partial^{r_U+r_V-1}$ for the coefficients of the greatest common right divisor of U and V, then we have the equation

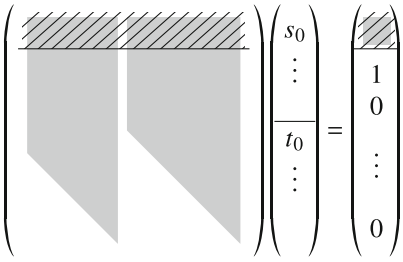
$$\operatorname{Syl}(U, V) \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{r_U-1} \\ t_0 \\ \vdots \\ t_{r_V-1} \end{pmatrix} = \begin{pmatrix} g_0 \\ g_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g_{r_U+r_V-1} \end{pmatrix}$$

where $\operatorname{Syl}(U, V) \in K^{(r_U+r_V)\times(r_U+r_V)}$ is the matrix whose first r_V columns are the coefficient vectors of $U, \partial U, \ldots, \partial^{r_V-1} U$ and whose last r_U columns are the coefficient vectors of $V, \partial V, \ldots, \partial^{r_U-1} V$. This matrix is called the *Sylvester matrix* for $U, V \in K[\partial]$. Its determinant is called the *resultant* of $U, V \in K[\partial]$ and is denoted by $\operatorname{res}(U, V)$. If the resultant is nonzero, we find a solution

$(s_0, \dots, s_{r_U-1}, t_0, \dots, t_{r_V-1})$ for any right hand side $(g_0, \dots, g_{r_U+r_V-1})$, in particular for the right hand side $(1, 0, \dots, 0)$. This means that in this case, we have $1 = SU + TV$ for some $S, T \in K[\partial]$, which implies that $\gcrd(U, V) = 1$, because every common right divisor of U and V must also be a right divisor of $SU + TV$, no matter what S and T are, and since 1 obviously does not have any nontrivial right divisor, U and V cannot have a nontrivial common right divisor. We say that they are *right coprime*.

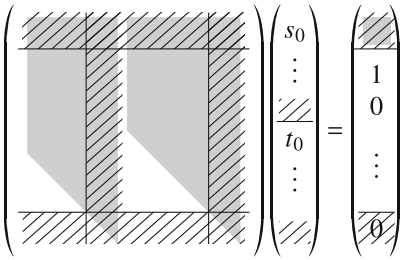
Conversely, if U and V are right coprime, then $SU + TV = 1$ for some $S, T \in K[\partial]$ with $\ord(S) < r_V$ and $\ord(T) < r_U$, and consequently $\partial^i SU + \partial^i TV = \partial^i$ for every $i \in \mathbb{N}$. From Theorem 4.28 below, it follows that there also are $S', T' \in K[\partial]$ with $\ord(S') \leq r_V$ and $S'U + T'V = 0$. Therefore, $\rrem(\partial^i S, S')U + (\partial^i T - \rquo(\partial^i S, S')T')V = \partial^i$ for every $i = 0, \dots, r_V + r_U - 1$. Because of $\ord(\rrem(\partial^i S, S')) < r_V$ and $\ord(\partial^i) < r_V + r_U$, we must have $\ord(\partial^i T - \rquo(\partial^i S, S')T') < r_U$ for all such i. This implies that the image of the Sylvester matrix contains all unit vectors in $K^{r_V+r_U}$, and thus consists of the full space. The kernel is therefore just $\{0\}$, and this in turn implies that the resultant is nonzero. We have thus shown that $\gcrd(U, V) = 1$ if and only if $\res(U, V) \neq 0$.

In the general case, suppose the greatest common right divisor of U and V has order r. Then in the linear system above we have $g_r = 1$ and $g_{r+1} = \cdots = g_{r_U+r_V-1} = 0$, and if we cut away the rows corresponding to $g_0, \dots, g_{r-1}$, we obtain the following linear system for the coefficients of S and T:



Every solution vector $(s_0, \dots, t_0, \dots) \in K^{r_u+r_v}$ of this system gives rise to operators $S, T \in K[\partial]$ such that $SU + TV = \partial^r + \cdots$. The right hand side must in fact be the greatest common right divisor of U and V, because the greatest common right divisor of U and V must be a right divisor of the left hand side, hence also of the right side, and since both the gcrd and the right hand side have order r, they must be equal.

By part 3 of Theorem 4.21 we also know that we may assume $s_{r_V-i} = 0$ for $i = 1, \dots, r$ and $t_{r_U-i} = 0$ for $i = 1, \dots, r$, so we can remove these variables from the linear system and delete the corresponding columns from the Sylvester matrix. Since $[\partial^k]\partial^i U = 0$ for $k > i + r_U$ and $[\partial^k]\partial^i V = 0$ for $k > i + r_V$, the Sylvester matrix contains zeros in the areas indicated by the unshaded triangular regions in the figure above. Removing the $2r$ unnecessary columns will therefore leave us with a matrix that has at least r rows that only contain zeros, and we may drop these as well. The situation then looks as follows:



As a result of this discussion, we get an alternative algorithm for computing $\operatorname{gcrd}(U, V)$: solve the above linear system in turn for $r = 0, 1, 2, \ldots$. The first r for which the system has a solution gives us the cofactors S, T from which we can get the final result via $\operatorname{gcrd}(U, V) = SU + TV$. This algorithm has the nice feature that it gives us access to a bound on the size of the coefficients that may appear in the gcrd.

Theorem 4.23 *Suppose that $K = C(x)$ and that $\sigma, \delta \colon K \to K$ map polynomials to polynomials and do not increase degrees, i.e.,* $\deg(\sigma(p)), \deg(\delta(p)) \leq \deg(p)$ *for all $p \in C[x]$. Let $U, V \in C[x][\partial] \subseteq K[\partial]$ be such that $U, V \neq 0$ and* $\operatorname{lc}(V)U \neq \operatorname{lc}(U)V$. *Let $r_U = \operatorname{ord}(U)$, $r_V = \operatorname{ord}(V)$, and let $d_U, d_V \in \mathbb{N}$ be such that all coefficients of U have degree at most d_U and all coefficients of V have degree at most d_V. Let $G = \operatorname{gcrd}(U, V) \in K[\partial]$ and $r = \operatorname{ord}(G)$. Then the coefficients of G are elements of $C(x)$ whose numerators and denominators are polynomials of degree at most $(r_V - r)d_U + (r_U - r)d_V$.* □

Proof Continuing the preceding discussion, we find that the combined coefficient vector

$$(s_0, \ldots, s_{r_V - r - 1}, t_0, \ldots, t_{r_U - r - 1})$$

of $S, T \in K[\partial]$ such that $\operatorname{gcrd}(U, V) = SU + TV$ is a solution of an inhomogeneous linear system obtained from the Sylvester matrix by deleting $2r$ rows and $2r$ columns. Because of the assumptions on σ and δ, we have $\deg([\partial^k]\partial^i U) \leq d_U$ and $\deg([\partial^k]\partial^i V) \leq d_V$ for all i and k, so that the matrix has $r_V - r$ columns with entries of degree at most d_U and $r_U - r$ columns with entries of degree at most d_V. By Cramer's rule, the denominator of the coefficients of S and T is the determinant of this matrix, i.e., a polynomial of degree at most $(r_V - r)d_U + (r_U - r)d_V$, the numerators of the coefficients of S are polynomials of degree at most $(r_V - r)d_U + (r_U - r)d_V - d_U$, and the numerators of the coefficients of T are polynomials of degree at most $(r_V - r)d_U + (r_U - r)d_V - d_V$. The announced bound for $G = SU + TV$ follows. ■

In the setting of this theorem, a greatest common right divisor can be computed in polynomial time, because it suffices to solve at most r_V linear systems of size

at most $(r_U + r_V) \times (r_U + r_V)$ with polynomial entries of degree at most d_U, d_V. A naive implementation of Algorithm 4.20 will be slower, because the coefficients of G, S, T can grow dramatically during the execution of the loop, even though Theorem 4.23 guarantees that the final result will be of reasonable size. What lets the size drop in the end is the multiplication with $\mathrm{lc}(G)^{-1}$ from the left, and for an implementation of Algorithm 4.20 we reiterate the advice to introduce such a normalization step in each iteration of the loop. In the case $K = C(x)$, an even more careful implementation will avoid working with rational functions and ensure that G belongs to $C[x][\partial]$ at all times, and that the coefficients of G are coprime (assuming, as usual, that they are written to the left of the powers of ∂).

The greatest common right divisor is useful for showing that every left ideal of $K[\partial]$ is generated by a single element. It is also useful for describing the intersection of two solution spaces. The details are as follows.

Theorem 4.24 *Let F be a $K[\partial]$-module and let $A, B \in K[\partial]$. Then*

1. $V(A) \cap V(B) = V(\mathrm{gcrd}(A, B))$.
2. $\langle A, B\rangle = \langle \mathrm{gcrd}(A, B)\rangle$. □

Proof

1. "$\subseteq$": If $f \in V(A) \cap V(B)$, then $A \cdot f = B \cdot f = 0$, and then $(SA + TB) \cdot f = 0$ for every $S, T \in K[\partial]$. Taking S, T appropriately, we find $\mathrm{gcrd}(A, B) \cdot f = 0$, so $f \in V(\mathrm{gcrd}(A, B))$.
   "$\supseteq$": Writing $G = \mathrm{gcrd}(A, B)$, we have $A = \tilde{A}G$ and $B = \tilde{B}G$ for certain $\tilde{A}, \tilde{B} \in K[\partial]$, so if $f \in V(\mathrm{gcrd}(A, B))$, then $G \cdot f = 0$ implies $A \cdot f = (\tilde{A}G) \cdot f = \tilde{A} \cdot (G \cdot f) = \tilde{A} \cdot 0 = 0$ and similarly, $B \cdot f = (\tilde{B}G) \cdot f = 0$, so $f \in V(A) \cap V(B)$.
2. "$\subseteq$": With $G = \mathrm{gcrd}(A, B)$, we have $A = \tilde{A}G$ and $B = \tilde{B}G$ for certain $\tilde{A}, \tilde{B} \in K[\partial]$. Every $P \in \langle A, B\rangle$ can be written as $P = UA + VB$ for certain $U, V \in K[\partial]$, and $P = UA + VB = (U\tilde{A} + V\tilde{B})G$ shows $P \in \langle G\rangle$.
   "$\supseteq$": With $G = \mathrm{gcrd}(A, B)$ and $S, T \in K[\partial]$ such that $G = SA + TV$, every element $P \in \langle G\rangle$, say $P = \tilde{P}G$ for some $\tilde{P} \in K[\partial]$, can be written as $P = \tilde{P}SA + \tilde{P}SB$ and therefore belongs to $\langle A, B\rangle$. ■

Let us now turn from the greatest common right divisor to the least common left multiple. In the commutative case, the greatest common divisor and the least common left multiple are related through the formula $\mathrm{lc}(p)\,\mathrm{lc}(q)\,\mathrm{gcd}(p, q)\,\mathrm{lcm}(p, q) = pq$ (Exercise 11), which holds for all $p, q \in C[x]$ and allows us to compute either one of $\mathrm{gcd}(p, q), \mathrm{lcm}(p, q)$ if we know how to compute the other. Unfortunately, the formula does not hold in the general Ore setting. For example, for $U = xD - 1$ and $V = D + 1$, we have $UV = xD^2 + (x - 1)D - 1$, $\mathrm{gcrd}(U, V) = 1$, and $\mathrm{lclm}(U, V) = (x + 1)D^2 + xD - 1$. Also, $VU = xD^2 + xD - 1$ does not match.

Before we discuss the computation of least common left multiples, observe that the least common left multiple of $U, V \in K[\partial]$ in the sense of Definition 4.19 is at the same time a common left multiple of minimal order. For if $P \in K[\partial] \setminus \{0\}$ is a common left multiple of U, V of minimal order and $P' \in K[\partial]$ is another common

left multiple of U, V, then $\operatorname{rrem}(P', P) = P'P - \operatorname{rquo}(P', P)P$ is also a common left multiple of U and V, and since $\operatorname{ord}(\operatorname{rrem}(P', P)) < \operatorname{ord}(P)$ and $\operatorname{ord}(P)$ is minimal, we must have $\operatorname{rrem}(P', P) = 0$, which is exactly the condition for P to be a right divisor of P. With this knowledge, we can prove the following theorem, which contains some counterparts of the previous theorem and reveals that we have met the least common left multiple already in the previous section.

Theorem 4.25 *Let F be a $K[\partial]$-module and let $A, B \in K[\partial]$. Then*

1. $V(A) + V(B) \subseteq V(\operatorname{lclm}(A, B))$.
2. $\langle A\rangle \cap \langle B\rangle = \langle \operatorname{lclm}(A, B)\rangle$.
3. $\operatorname{lclm}(A, B) = A \oplus B$. □

Proof

1. Writing $m = \operatorname{lclm}(u, v)$, we have $m = \tilde{u}u = \tilde{v}v$ for certain $\tilde{u}, \tilde{v} \in K[\partial]$, so if $f \in V(u) + V(v)$, say $f = f_u + f_v$ for some $f_u \in V(u)$ and some $f_v \in V(v)$, then $m \cdot f = m \cdot (f_u + f_v) = (m \cdot f_u) + (m \cdot f_v) = (\tilde{u}u \cdot f_u) + (\tilde{v}v \cdot f_v) = 0$, so $f \in V(m)$.
2. "$\subseteq$": If $P \in \langle A\rangle \cap \langle B\rangle$, then $P = \tilde{A}A = \tilde{B}B$ for certain $\tilde{A}, \tilde{B} \in K[\partial]$, so P is a common left multiple of A and B, and therefore a left multiple of $\operatorname{lclm}(A, B)$, and therefore an element of $\langle \operatorname{lclm}(A, B)\rangle$.
   "$\supseteq$": If $M = \operatorname{lclm}(A, B)$, then $M = \tilde{A}A = \tilde{B}B$ for certain $\tilde{A}, \tilde{B} \in K[\partial]$, and if $P \in \langle \operatorname{lclm}(A, B)\rangle$, then $P = \tilde{M}M$ for some $\tilde{M} \in K[\partial]$, and $P = \tilde{M}\tilde{A}A = \tilde{M}\tilde{B}B$ shows that $P \in \langle A\rangle \cap \langle B\rangle$.
3. Recall that $A \oplus B$ was defined as the monic minimal annihilating operator of $([1], [1]) \in K[\partial]/\langle A\rangle \times K[\partial]/\langle B\rangle$. With the definition of the action of $K[\partial]$ on this module, we have $M \cdot ([1], [1]) = ([M], [M]) = ([0], [0])$ if and only if $M \in \langle A\rangle \cap \langle B\rangle$. It follows from the previous part that $\langle \operatorname{lclm}(A, B)\rangle$ is the annihilator of $([1], [1])$. Since $\operatorname{lclm}(A, B)$ is the unique monic element of minimal order in this ideal, it must be equal to $A \oplus B$. ■

Observe that only one inclusion is claimed in part 1. The other inclusion is false in general. Counterexamples can be constructed from elements of $K[\partial]$ whose solution space in F does not have the largest possible dimension.

Example 4.26 Consider the action of $C(x)[D]$ on $F = C(x)$ and let $A = xD^2 + D$ and $B = x(x + 1)D^2 + D$. We then have $\operatorname{lclm}(A, B) = xD^3 + 2D^2$, and it can be checked that both $V(A)$ and $V(B)$ are C-vector spaces generated by 1, so $V(A) + V(B) = V(A) = V(B)$. However, $V(\operatorname{lclm}(A, B))$ contains the additional polynomial $x \notin V(A) + V(B)$.

It becomes more clear what is going on if we replace F by a larger differential field. In $F = C(x, \log(x))$, the vector space $V(A)$ is generated by 1 and $\log(x)$, and the vector space $V(B)$ is generated by 1 and $x + \log(x)$. We see that in this case, $V(A) + V(B)$ contains the missing solution x of $\operatorname{lclm}(A, B)$. □

Part 3 of Theorem 4.25 motivates the following algorithm for computing the least common left multiple of any two elements of $K[\partial]$.

Algorithm 4.27 *(Least common left multiple)*
Input: $U, V \in K[\partial]$.
Output: $\operatorname{lclm}(U, V) \in K[\partial]$.

1 *if* $U = 0$ *or* $V = 0$ *then*
2 *Return* 0.
3 *Set* $U_0 = V_0 = 1 \in K[\partial]$.
4 *for* $r = 1, 2, \ldots,$ *do*
5 *Compute* $U_r = \operatorname{rrem}(\partial U_{r-1}, U) \in K[\partial]$.
6 *Compute* $V_r = \operatorname{rrem}(\partial V_{r-1}, V) \in K[\partial]$.
7 *Check whether there exists* $(p_0, \ldots, p_r) \in K^{r+1} \setminus \{0\}$ *with*

$$p_0 U_0 + \cdots + p_r U_r = p_0 V_0 + \cdots + p_r V_r = 0.$$

8 *if yes then*
9 *Return* $\frac{p_0}{p_r} + \frac{p_1}{p_r}\partial + \cdots + \frac{p_{r-1}}{p_r}\partial^{r-1} + \partial^r$.

Theorem 4.28 *Algorithm 4.27 is correct and terminates. In particular, we have*

$$\operatorname{ord}(\operatorname{lclm}(U, V)) \leq \operatorname{ord}(U) + \operatorname{ord}(V)$$

for all $U, V \in K[\partial] \setminus \{0\}$. □

Proof First note that for all $i \in \mathbb{N}$ we have $U_i = \operatorname{rrem}(\partial^i, U)$ and $V_i = \operatorname{rrem}(\partial^i, V)$, so that line 7 ensures that

$$\operatorname{rrem}(p_0 + p_1\partial + \cdots + p_r\partial^r, U) = \operatorname{rrem}(p_0 + p_1\partial + \cdots + p_r\partial^r, V) = 0,$$

which means that $P = p_0 + p_1\partial + \cdots + p_r\partial^r$ is a common left multiple of U and V. If we find a nonzero coefficient vector $(p_0, \ldots, p_r)$ in this step, we must have $p_r \neq 0$, for otherwise we would have found the solution already in an earlier iteration. It is therefore safe to divide by p_r (from the left) and to return $p_r^{-1}P$ as the correct result.

For the termination, observe that every U_i is a K-linear combination of the powers $1, \partial, \ldots, \partial^{\operatorname{ord}(U)-1}$ and that each V_i is a K-linear combination of $1, \partial, \ldots, \partial^{\operatorname{ord}(V)-1}$. Therefore, the coefficient comparison with respect to powers of ∂ done in line 7 leads to a linear system with $\operatorname{ord}(U) + \operatorname{ord}(V)$ equations and $r + 1$ equations, which must have a solution as soon as $r > \operatorname{ord}(U) + \operatorname{ord}(V)$. This proves termination as well as the claimed bound on the order of the output. ■

Example 4.29

1. In $C(x)[D]$, consider $U = (2x+3)xD^2 + 2(4x^2+3x-3)D + 2(4x^2-3)$, $V = (x+1)(x-1)D^2 + 2(2x^2-x-2)D + 2(2x^2-2x-1)$. With $U_0 = V_0 = 1$ and $U_1 = V_1 = D$ we obviously have no nontrivial solution yet. In the next step, we get

$$U_2 = \operatorname{rrem}(D^2, U) = \frac{2(4x^2+3x-3)}{x(2x+3)}D - \frac{2(4x^2-3)}{x(2x+3)},$$
$$V_2 = \operatorname{rrem}(D^2, V) = \frac{2(2x^2-x-2)}{(x+1)(x-1)}D - \frac{2(2x^2-2x-1)}{(x+1)(x-1)},$$

and there could be $p_0, p_1, p_2 \in K$ such that $p_0U_0 + p_1U_1 + p_2U_2 = p_0V_0 + p_1V_1 + p_2V_2 = 0$. Coefficient comparison leads to the linear system

$$\begin{pmatrix} 1 & 0 & -\frac{2(4x^2-3)}{x(2x+3)} \\ 0 & 1 & \frac{2(4x^2+3x-3)}{x(2x+3)} \\ 1 & 0 & -\frac{2(2x^2-2x-1)}{(x+1)(x-1)} \\ 0 & 1 & \frac{2(2x^2-x-2)}{(x+1)(x-1)} \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} = 0,$$

whose only solution is zero. We continue with

$$U_3 = \frac{6(x-1)(4x^2-4x-1)}{x^2(2x+3)}D + \frac{2(16x^3-12x^2-12x+3)}{x^2(2x+3)},$$
$$V_3 = \frac{12(x^2-x-1)}{(x+1)(x-1)}D + \frac{4(4x^2-6x-1)}{(x+1)(x-1)},$$

but will still not find a solution. In the next step, we have

$$U_4 = \frac{16(4x^3-3x^2-6x+3)}{x^2(2x+3)}D - \frac{48(2x-1)(x^2-x-1)}{x^2(2x+3)},$$
$$V_4 = \frac{16(2x+1)(x-2)}{(x+1)(x-1)}D - \frac{48x(x-2)}{(x+1)(x-1)},$$

and the corresponding linear system

$$\begin{pmatrix} 1 & 0 & -\frac{2(4x^2-3)}{x(2x+3)} & \frac{2(16x^3-12x^2-12x+3)}{x^2(2x+3)} & -\frac{48(2x-1)(x^2-x-1)}{x^2(2x+3)} \\ 0 & 1 & \frac{2(4x^2+3x-3)}{x(2x+3)} & \frac{6(x-1)(4x^2-4x-1)}{x^2(2x+3)} & \frac{16(4x^3-3x^2-6x+3)}{x^2(2x+3)} \\ 1 & 0 & -\frac{2(2x^2-2x-1)}{(x+1)(x-1)} & \frac{4(4x^2-6x-1)}{(x+1)(x-1)} & -\frac{48x(x-2)}{(x+1)(x-1)} \\ 0 & 1 & \frac{2(2x^2-x-2)}{(x+1)(x-1)} & \frac{12(x^2-x-1)}{(x+1)(x-1)} & \frac{16(2x+1)(x-2)}{(x+1)(x-1)} \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = 0$$

must have a nonzero solution because it has more variables than equations. A basis vector of the solution space translates into the operator

$$\operatorname{lclm}(U, V) = D^4 + 8D^3 + 24D^2 + 32D + 16.$$

This is really a common left multiple of U and V because we have

$$\begin{aligned}\operatorname{lclm}(U,V) &= \left(\frac{1}{x(2x+3)}D^2 + \frac{2(4x+5)}{x(2x+3)^2}D + \frac{8(x+1)}{x(2x+3)^2}\right)U \\ &= \left(\frac{1}{(x+1)(x-1)}D^2 + \frac{2(2x^2-x-2)}{(x+1)^2(x-1)^2}D + \frac{4(x^2-x-1)}{(x+1)^2(x-1)^2}\right)V.\end{aligned}$$

2. In $C(x)[S]$, consider $U = (x+2)(2x-1)S^2 - 8(x^2+3x-1)S + 4(x+4)(2x+1)$ and $V = (x+4)(x+3)S^2 - 2(x+5)(2x+5)S + 4(x+4)(x+5)$. An analogous computation as above yields

$$\operatorname{lclm}(U,V) = S^3 - \frac{2(3x+10)}{x+3}S^2 + \frac{4(3x+11)}{x+3}S - \frac{8(x+4)}{x+3},$$

and this is really a common left multiple because it can be written as

$$\begin{aligned}\operatorname{lclm}(U,V) &= \left(\frac{1}{(2x+1)(x+3)}S - \frac{2}{(2x+1)(x+3)}\right)U \\ &= \left(\frac{1}{(x+5)(x+3)}S - \frac{2}{(x+5)(x+3)}\right)V.\end{aligned}$$

□

Again, we observe a phenomenon that cannot happen in the commutative case: the coefficients of the least common left multiple are smaller than those of the input. This does not happen generically, but as the example shows, it can happen. What happens generically though is that higher order common left multiples may have lower degree than the least common left multiple.

Theorem 4.30 *Suppose that $K = C(x)$ and that $\sigma, \delta\colon K \to K$ map polynomials to polynomials and do not increase degrees, i.e., $\deg(\sigma(p)), \deg(\delta(p)) \leq \deg(p)$ for all $p \in C[x]$. Let $U, V \in C[x][\partial] \subseteq K[\partial]$, $r_U = \operatorname{ord}(U)$, $r_V = \operatorname{ord}(V)$, and let $d_U, d_V \in \mathbb{N}$ be such that all coefficients of U have degree at most d_U and all coefficients of V have degree at most d_V.*

1. *Let $P = \operatorname{lclm}(U,V) \in K[\partial]$ and $r = \operatorname{ord}(P)$. Then the coefficients of P are elements of $C(x)$ whose numerators and denominators are polynomials of degree at most*

$$(r+1-r_V)d_U + (r+1-r_U)d_V.$$

2. *For every $r \geq r_U + r_V$ and every*

$$d > d_U + d_V - 1 + \frac{r_V d_U + r_U d_V}{r - r_U - r_V + 1}$$

there exists a common left multiple of U and V of order r with polynomial coefficients of degree at most d. □

Proof

1. If $S, T \in K[\partial]$ are such that $P = SU = TV$, then $\operatorname{ord}(S) = r - r_U$ and $\operatorname{ord}(T) = r - r_V$. Consider an ansatz

$$(s_0 + s_1\partial + \cdots + s_{r-r_U}\partial^{r-r_U})U - (t_0 + t_1\partial + \cdots + t_{r-r_V}\partial^{r-r_V})V = 0$$

with undetermined coefficients $s_0, \ldots, s_{r-r_U}$ and $t_0, \ldots, t_{r-r_V}$. Equating coefficients of ∂^i to zero, for $i = 0, \ldots, r$, leads to a linear system with $(r - r_U + 1) + (r - r_V + 1)$ variables and $r + 1$ equations. By assumption on σ and δ, the corresponding matrix has $r - r_U + 1$ columns with entries of degree at most d_U and $r - r_V + 1$ columns with entries of degree at most d_V. Because of the uniqueness of $\operatorname{lclm}(U, V)$, the linear system has a solution space in $C(x)^{2r+2-r_U-r_V}$ of dimension 1, so the corresponding matrix has rank $2r + 1 - r_U - r_V$. By Theorem 1.29, the solution space is generated by a vector $(s_0, \ldots, s_{r-r_U}, t_0, \ldots, t_{r-r_V}) \in C[x]^{2r+2-r_U-r_V}$ with

$$\begin{aligned}\deg(s_i) &\le (r + 1 - r_V)d_U + (r + 1 - r_U)d_V - d_U \quad \text{and}\\ \deg(t_j) &\le (r + 1 - r_V)d_U + (r + 1 - r_U)d_V - d_V\end{aligned}$$

for all $i = 0, \ldots, r - r_U$ and $j = 0, \ldots, r - r_V$. The announced degree bound for $P = SU = TV$ follows.

2. Let $r \ge r_U + r_V$ and $d > d_U + d_V - 1 + \frac{r_V d_U + r_U d_V}{r - r_U - r_V + 1}$ and make an ansatz

$$S = \sum_{i=0}^{r-r_U} \sum_{j=0}^{d-d_U} s_{i,j} x^j \partial^i \quad T = \sum_{i=0}^{r-r_V} \sum_{j=0}^{d-d_V} t_{i,j} x^j \partial^i$$

with undetermined coefficients $s_{i,j}, t_{i,j} \in C$. We show that these coefficients can be instantiated such that $SU = TV$. Indeed, equating the coefficients of $SU - TV$ with respect to $x^j\partial^i$ to zero gives a C-linear system with $(r - r_U + 1)(d - d_U + 1) + (r - r_V + 1)(d - d_V + 1)$ variables and no more than $(r+1)(d+1)$ equations. Because of the assumption on d, we have

$$\begin{aligned}&(r - r_U + 1)(d - d_U + 1) + (r - r_V + 1)(d - d_V + 1) - (r + 1)(d + 1)\\ &= (r + 1 - r_U - r_V)(d + 1 - d_U - d_V) - r_V d_U - r_U d_V\\ &> (r + 1 - r_U - r_V)\Big(d_U + d_V - 1 + \frac{r_V d_U + r_U d_V}{r - r_U - r_V + 1} + 1 - d_U - d_V\Big)\\ &\qquad - r_V d_U - r_U d_V\\ &= 0\end{aligned}$$

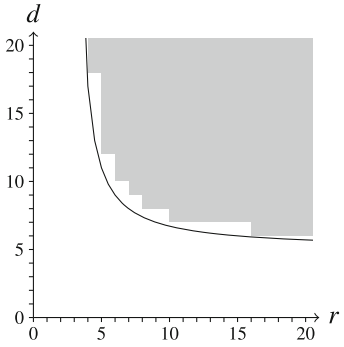
and therefore more variables than equations. The nontrivial solution gives rise to a common left multiple SU whose order may still be less than r, because some of

the coefficients $s_{i,j}$ of the nonzero solution may be zero. But since also $\partial^i SU$ is a common left multiple for every choice of i, we can get a left multiple of order exactly r with coefficients of degree at most d. ■

Example 4.31 Consider the operators

$$U = (2x^3 + 2x^2 + 8)D^2 + (7x^2 + 5x + 3)D + (3x^3 + x^2 + x + 7),$$
$$V = (9x^3 + 8x^2 + 6x + 3)D^2 + (3x^3 + 2x^2 + 7x + 5)D + (4x^3 + x + 9).$$

In the following figure, the gray region marks all the points (r, d) for which there exists a common left multiple of U and V of order r and degree d. By the theorem above, all points $(r, d) \in \mathbb{N}^2$ satisfying $d \geq 5 + \frac{12}{r-3}$ belong to this gray region. As the white space between the curve and the gray region contains no points with integer coordinates, the bounds provided by the theorem are sharp in this example.



□

Part 2 of Theorem 4.30 allows us to get common left multiples of smaller degree if we allow for larger orders. If we allow a very large order, we can get the degree down to $d_U + d_V$. In general, there will not exist a common left multiple of lower degree. Just consider two differential operators $U, V \in C[x][D]$ whose leading coefficients are squarefree coprime polynomials of degree d_U and d_V, respectively, with roots that are non-apparent singularities in the sense of Definition 3.17. Since every solution of U or V must also be a solution of any common left multiple of U and V, any such left multiple must have a leading coefficient that is a multiple of $\operatorname{lcm}(\operatorname{lc}(U), \operatorname{lc}(V))$, and if $\operatorname{lc}(U)$ and $\operatorname{lc}(V)$ are coprime as elements of $C[x]$, the degree of $\operatorname{lcm}(\operatorname{lc}(U), \operatorname{lc}(V)) = \operatorname{lc}(U)\operatorname{lc}(V)$ is $d_U + d_V$.

If there are apparent singularities, the degrees of the left multiples may be smaller than the bound of Theorem 4.30. In fact, we can use left multiples to lift the discussion of apparent singularities of Sect. 3.2 to arbitrary Ore algebras. We do not even need to refer to solutions of an operator.

Definition 4.32 Let $P \in C[x][\partial]$, $r = \operatorname{ord}(P)$, and let $p \in C[x]$ be a factor of $\operatorname{lc}(P)$. Let $n \in \mathbb{N}$. We say that p is *removable* from P at cost n if there exists an

operator $Q \in C(x)[\partial]$ such that $QP \in C[x][\partial]$ and $\mathrm{lc}(QP) \mid \sigma^n(\mathrm{lc}(P)/p)$. In this case, we say that Q is a *p-removing* operator for P. □

For example, if $P = xD - 5 \in C[x][D]$, the factor $p = x$ is removable at cost $n = 5$ because for $Q = \frac{1}{x}D^5$ we have $QP = D^6$. This is in line with the examples we discussed in Sect. 3.2, but removability as defined above is not restricted to the differential case. For example, if $P = xS - (x+3) \in C[x][S]$, the factor $p = x$ is removable at cost $n = 3$, because for $Q = \frac{1}{x+3}(S-1)^3$ we have $QP = S^4 - 4S^3 + 6S^2 - 4S + 1$. One application of removing factors is that with a bit of luck it may allow us to show that a D-finite sequence has only integer terms.

Example 4.33

1. Consider the D-finite sequence $(a_n)_{n=0}^{\infty}$ defined by $a_0 = 2$, $a_1 = 3$, $a_3 = 14$, and

$$(n-1)a_{n+2} = (n^2 + 3n - 2)a_{n+1} - 2n(n+1)a_n \qquad (n \in \mathbb{N}).$$

   Computing the nth term of the sequence with this recurrence requires a division by $n-3$, and there is no obvious reason why this division should always produce integers. But in fact it does, because $x+1$ is removable for the operator $P = (x-1)S^2 - (x^2 + 3x - 2)S + 2x(x+1) \in C[x][S]$. More precisely, we have

$$\tfrac{1}{x}(S-2)P = S^3 - (x+7)S^2 + 4(x+3)S - 4(x+1),$$

   so the sequence $(a_n)_{n=0}^{\infty}$ also satisfies the recurrence

$$a_{n+3} = (n+7)a_{n+2} - 4(n+3)a_{n+1} - 4(n+1)a_n,$$

   from which it can easily be seen that $a_n \in \mathbb{Z}$ for all $n \in \mathbb{N}$.
2. An operator whose leading coefficient has non-removable factors may nevertheless have integer sequence solutions. For example $P = (x+2)S - (4x+2) \in C[x][S]$ is an annihilating operator for the sequence $(C_n)_{n=0}^{\infty}$ of Catalan numbers, and while we clearly have $C_n \in \mathbb{Z}$ for all $n \in \mathbb{N}$, the factor $x+2$ is not removable. □

If $p \in C[x]$ is irreducible and $k \in \mathbb{N}$, then according to Exercise 18 a p^k-removing operator can be assumed to be of the form

$$Q = \frac{1}{\sigma^n(p)^{e_n}}\partial^n + \frac{q_{n-1}}{\sigma^n(p)^{e_{n-1}}}\partial^{n-1} + \cdots + \frac{q_0}{\sigma^n(p)^{e_0}}$$

for some $e_0, \ldots, e_n \in \mathbb{N}$ and some $q_0, \ldots, q_{n-1} \in C[x]$ with $\deg(q_i) < e_i \deg(p)$ $(i = 0, \ldots, n-1)$. If someone gives us suitable n and $e_0, \ldots, e_n$, we can find $q_0, \ldots, q_{n-1}$ by making an ansatz $q_i = \sum_{j=0}^{e_i \deg(p)-1} q_{i,j}x^j$ with undetermined coefficients $q_{i,j}$, computing QP, and forcing its coefficients to be polynomials. A priori, the coefficients of QP are rational functions whose numerators can be written

as a linear combination of the unknown coefficients, and they become polynomials whenever the numerators are multiples of the denominators. We can enforce this by computing the remainders of all numerators with respect to the corresponding denominators and equate their coefficients to zero. This gives an inhomogeneous linear system for the unknowns $q_{i,j}$, which may or may not have a solution. If it has a solution and we instantiate $q_{i,j}$ accordingly, then Q is a p^k-removing operator for P.

The choice of n and $e_0, \ldots, e_n$ depends on the particular Ore algebra at hand. For the differential case, desingularization is covered in Sect. 3.3. For other algebras, bounds on n and $e_0, \ldots, e_n$ can be found in the literature.

There is another, more pragmatic, way to remove removable factors by computing a least common left multiple. In the proof of Theorem 4.30, we used an ansatz

$$(s_0 + s_1\partial + \cdots + s_{r_V}\partial^{r_V})U = (t_0 + t_1\partial + \cdots + t_{r_U}\partial^{r_U})V$$

and compared coefficients to obtain a linear system of equations whose solution vectors gave rise to the coefficients of operators S, T that we can multiply from the left to U, V, respectively, to obtain a common left multiple of U and V. We have some freedom to modify this ansatz. Suppose, for example, that $\operatorname{lc}(U)$ contains an irreducible factor $p \in C[x]$ which is removable at cost n, and that $Q \in C(x)[\partial]$ is a p-removing operator of order n. Consider the alternative ansatz

$$(s_0 + s_1\partial + \cdots + s_{n-1}\partial^{n-1} + s_n Q)U = (t_0 + t_1\partial + \cdots + t_{r_U}\partial^{r_U})V.$$

When $(s_0, \ldots, s_n, t_0, \ldots, t_{r_U}) \in C[x]^{n+r_U+2}$ is a nonzero solution vector of the resulting linear system, then $(s_0 + s_1\partial + \cdots + s_{n-1}\partial^{n-1} + s_n Q)U$ is a left multiple of U whose leading coefficient is $s_n \operatorname{lc}(QU)$, which is not a multiple of $\sigma^n(p)$ unless $\sigma^n(p)$ happens to be a factor of s_n. It can be shown that only for very few choices of V will it happen that $\sigma^n(p) \mid s_n$, so if we randomly choose an operator $V \in C[x][\partial]$ of order n, we can expect that left-multiplying $\operatorname{lclm}(U, V) \in C(x)[\partial]$ with the common denominator of its coefficients yields a left-multiple of U which lives in $C[x][\partial]$ and whose leading coefficient does not contain $\sigma^n(p)$ as a factor.

Note that although we assumed the knowledge about a p-removing operator Q in the discussion above, we do not actually need to know Q for computing $\operatorname{lclm}(U, V)$.

Note also that taking the least common left multiple of U with some other operator V is not only likely to remove removable factors, but it is also likely to introduce new factors. In general, the factor s_n of the new leading coefficient is not just a constant. In order to get a left-multiple of U with a smaller leading coefficient, we compute $g = \gcd(\operatorname{lc}(P), \sigma^n(\operatorname{lc}(U)))$ and $s, t \in C[x]$ with $g = s\operatorname{lc}(P) + t\sigma^n(\operatorname{lc}(U))$. Then $sP + t\partial^n U$ is a left-multiple of U whose leading coefficient g contains neither $\sigma^n(p)$ nor the factors that have been introduced by the lclm-computation.

Exercises

1⋆. Let $A, B, C \in K[\partial]$, $C \neq 0$. Prove or disprove:
- **a.** $\operatorname{rrem}(A, C) + \operatorname{rrem}(B, C) = \operatorname{rrem}(A + B, C)$.
- **b.** $\operatorname{rrem}(A\operatorname{rrem}(B, C), C) = \operatorname{rrem}(AB, C)$.
- **c.** $\operatorname{rrem}(\operatorname{rrem}(A, C)B, C) = \operatorname{rrem}(AB, C)$.

2. Show that for any $U, V \in K[\partial]$, there can be at most one greatest common right divisor and at most one least common left multiple.

3. Show that $\operatorname{gcrd}(U, \operatorname{gcrd}(V, W)) = \operatorname{gcrd}(\operatorname{gcrd}(U, V), W)$ for all $U, V, W \in K[\partial]$.

4. Compute $\operatorname{gcrd}(U, V)$ for $U = (x+1)\partial^2 + (x^2+2x-1)\partial - x^2(x+1)$, $V = (x+1)\partial^2 + (3x+1)\partial - 2x(x+1) \in C(x)[\partial]$ with $\sigma : C(x) \to C(x)$ defined by $\sigma(p(x)) = p(\frac{1-x}{1+x})$ for $p(x) \in C(x)$ and $\delta = 0$.

5. Compute S, T such that $\operatorname{gcrd}(U, V) = SU + TV$ for the two pairs of operators U, V considered in Example 4.22.

6⋆. Show part 4 of Theorem 4.21. *Hint:* First consider the case where U, V are coprime.

7. Let $U, V \in K[\partial] \setminus \{0\}$ and $G = \operatorname{gcrd}(U, V)$. Show that $S, T \in K[\partial]$ with $G = SU + TV$ and $\operatorname{ord}(S) < \operatorname{ord}(V) - \operatorname{ord}(G)$ and $\operatorname{ord}(T) < \operatorname{ord}(U) - \operatorname{ord}(G)$ do not exist if $U = cV$ for some $c \in K \setminus \{0\}$.

8⋆. Let $\sigma, \delta : C[x] \to C[x]$ be such that $\deg(\sigma(p)), \deg(\delta(p)) < \deg(p)$ for all $p \in C[x]$. Suppose that the application of σ or δ to a given polynomial $p \in C[x]$ of degree at most d costs $\mathrm{O}^{\sim}(d)$ operations in C. Let $r, d \in N$, and let $U, V \in C[x][\partial]$ with $\operatorname{ord}(V) < \operatorname{ord}(U) \leq r$ and with coefficients of degree at most d. Show that $\operatorname{gcrd}(U, V) \in C(x)[\partial]$ can be computed using $\mathrm{O}^{\sim}(r^{\omega}d)$ operations in C.

9. Find all $\alpha \in C$ for which $U = S^2 + (1 + 2x - x^2)S - x^2(x - \alpha)$ and $V = S^2 + \alpha S - x(x + \alpha - 1)$ have a nontrivial greatest common right divisor.

10⋆. A consequence of Theorem 4.24 is that $K[\partial]$ is a principle left ideal domain, i.e., every left ideal of $K[\partial]$ is generated by a single element. In contrast, show that $C[x][\partial]$ is in general not a principle left ideal domain.

11⋆. Show that $\operatorname{lc}(p)\operatorname{lc}(q)\operatorname{gcd}(p, q)\operatorname{lcm}(p, q) = pq$ for all $p, q \in C[x]$. Where do you need commutativity?

12⋆. Let $a, b \in C(x) \setminus \{0\}$ and consider an algebraic function y with minimal polynomial $y^2 + ay + b$. Let $L \in C(x)[D]$ be the monic minimal order annihilating operator of y (cf. Theorem 3.29). Show that L is a least common left multiple of two first order operators.

13⋆. Show that we have $\operatorname{ord}(U) + \operatorname{ord}(V) = \operatorname{ord}(\operatorname{gcrd}(U, V)) + \operatorname{ord}(\operatorname{lclm}(U, V))$ for all $U, V \in K[\partial] \setminus \{0\}$.

14⋆. When the extended Euclidean algorithm terminates, the components of (G, S, T, G', S', T') are such that $G = \operatorname{gcrd}(U, V) = SU + TV$ and $G' = 0$. Show that furthermore, $\operatorname{lclm}(U, V) = aS'U = bT'V$ for certain nonzero $a, b \in K$.

Hint: Consider the $K[\partial]$-submodule of $K[\partial]^3$ generated by $(U, 1, 0)$ and $(V, 0, 1)$.

15. Compute $\operatorname{lclm}(U, V)$ for U, V from Exercise 4, for the following settings:

a. $\sigma(p(x)) = p(\frac{x-1}{x+1})$ for all $p \in C(x)$, and $\delta = 0$;

b. $\sigma = \mathrm{id}$ and $\delta = 0$;

c. $\sigma = \mathrm{id}$ and $\delta(p(x)) = \frac{x-1}{x+1} p'(x)$ for all $p \in C(x)$.

16. Find $A, B \in C(x)[S]$ and a $C(x)[S]$-module F such that $V(A) + V(B) \subsetneq V(\operatorname{lclm}(A, B))$ in F (cf. Example 4.26).

17⋆⋆. Let $\sigma, \delta \colon C[x] \to C[x]$ be as in Exercise 8. Let $r \in N, d \in \mathbb{N}$ and $U, V \in C[x][\partial]$ with $\operatorname{ord}(U), \operatorname{ord}(V) \le r$ and coefficients of degree at most d. Show that computing $\operatorname{lclm}(U, V)$ costs no more than $\mathrm{O}^{\sim}(r^{\omega} d)$ operations in C.

Hint: Analyze the algorithm implicit in the proof of Theorem 4.30.

18⋆. Let $P \in C[x][\partial]$ and let $p \in C[x]$ be an irreducible factor of $\operatorname{lc}(P)$ such that p^k is removable at cost n from P, for some $k \in \mathbb{N}$. Show that there exist $e_0, \ldots, e_n \in \mathbb{N}$ and $q_0, \ldots, q_{n-1} \in C[x]$ with $\deg q_i < e_i \deg(p)$ $(i = 0, \ldots, n)$ such that

$$Q = \frac{1}{\sigma^n(p)^{e_n}} \partial^n + \frac{q_{n-1}}{\sigma^n(p)^{e_{n-1}}} \partial^{n-1} + \cdots + \frac{q_0}{\sigma^n(p)^{e_0}}$$

is a p^k-removing operator for P.

19⋆. Show that if $p_1, p_2 \in C[x]$ are removable at cost n from some operator $P \in C[x][\partial]$, then also $\operatorname{lcm}(p_1, p_2)$ is removable at cost n from P.

20. Analogous to right quotients, right remainders, right divisors, and left multiples, we can also define left quotients, left remainders, left divisors, and right multiples. Let $U = D^2 - (x^2 + 1)$, $V = xD^2 + (x^2 - 1)D - 2x \in C(x)[D]$.

a. Compute the greatest common left divisor of U and V.

b. Compute the least common right multiple of U and V.

21. Prove or disprove: $U, V \in K[\partial]$ have a nontrivial greatest common right divisor if and only if they have a nontrivial greatest common left divisor.

22. Can a recurrence have a d'Alembertian solution $\sum_{k=1}^{n} h_k$ for some hypergeometric term h_k without also having a hypergeometric solution similar to h_n?

References

Common right divisors and left multiples were already computed in the early days of differential and difference operators and noncommutative polynomial rings. Bostan,

Chyzak, Li, and Salvy [88] trace back the history deeply into 19th century, so that Ore's paper from 1933 [344] almost seems recent. Even more recent is the exposition in the tutorial paper of Bronstein and Petkovšek [115]. The paper [88] contains a careful comparison of various algorithms for computing common left multiples.

Although the Gauss lemma does not literally hold in the case of Ore algebras, there is a theorem due to Kovacic [294, Proposition 2] which can be viewed as a version of the statement for differential operators.

Surgery on the Sylvester matrix is known as subresultant theory and was introduced into computer algebra by Collins [163] for improving the computation of polynomial gcds in the commutative case. The theory was adapted to the case of Ore polynomials by Li [308, 309]. Jaroschek [248, 249] uses Li's subresultants to speed up the computation of gcrd's in $K[\partial]$. Grigoriev [226] points out that the idea of subresultants can be extended to the case of more than two operators, and proposes an algorithm for computing the greatest common right divisor of several differential operators in polynomial time.

Part 2 of Theorem 4.30 belongs to a family of results concerning the more general phenomenon that we can sometimes get lower degree coefficients by allowing an operator to have higher degree. The relationships between orders and degrees are expressed as order-degree curves. The phenomenon was observed for the first time by Bostan, Chyzak, Salvy, Lecerf, and Schost for differential equations of algebraic functions [83]. The result discussed in Theorem 4.30 is taken from a paper of Kauers [265], which also contains results for other closure properties. Order-degree curves in the context of summation and integration are discussed in Sect. 5.5. A connection between order-degree curves and removability of singularities was observed by Chen, Jaroschek, Kauers, and Singer [136].

Bounds on the order n and the exponents e_i of a p-removing operator have been derived for the shift case by Abramov, Barkatou and van Hoeij [9, 24], and for the q-shift case by Koutschan and Zhang [292]. The pragmatic way to remove removable factors was studied by Chen, Kauers, and Singer [141], although it had been used long before this paper in internal parts of the Maple library. A refined version of removability, which also allows to remove constant factors, was proposed by Zhang [473, 474].

4.3 Several Functions

We have seen in Exercise 26 of Sect. 4.1 that the solutions f of an inhomogeneous equation $P \cdot f = g$ must be D-finite as soon as the inhomogeneous part g is D-finite. In this case, we have an equation $Q \cdot g = 0$ for some nonzero operator Q, and we can view the two equations $P \cdot f - g = 0$, $Q \cdot g = 0$ as a coupled system of two equations for two unknown functions. Every pair (f, g) of functions that forms a solution of this system will be such that both f and g are D-finite.

In this section we consider coupled systems of functional equations more systematically. For a fixed Ore algebra $K[\partial]$ and a $K[\partial]$-module F, we consider equations of the form

$$I_r\partial^m \cdot f + A_{m-1}\partial^{m-1} \cdot f + \cdots + A_0 \cdot f = 0$$

where $f = (f_1, \ldots, f_r) \in F^r$ is a vector of unknown functions, ∂ is understood to act componentwise on such vectors, i.e., $\partial \cdot f = (\partial \cdot f_1, \ldots, \partial \cdot f_r)$, and $A_0, \ldots, A_{m-1}$ are given elements of $K^{r\times r}$.

The first observation about such equations is that it suffices to consider the case $m = 1$, because we can always reduce to this situation at the cost of increasing the size of the matrices. An element f of F^r is a solution of the above equation if and only if the element $\tilde{f} = (f, \partial \cdot f, \ldots, \partial^{m-1} \cdot f)^T$ of F^{mr} is a solution of the matrix equation

$$\partial \cdot \tilde{f} = \begin{pmatrix} 0 & I_r & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & I_r \\ -A_0 & \cdots & \cdots & \cdots & -A_{m-1} \end{pmatrix} \tilde{f}.$$

The more interesting question concerns the opposite direction: instead of lowering the order of the equation at the cost of increasing the size of the matrices, can we also decrease the size of the matrices at the cost of increasing the order of the equations? According to the following proposition, the answer is yes.

Proposition 4.34 *Let $K[\partial]$ be an Ore algebra and F be a $K[\partial]$-module. Let $A \in K^{r\times r}$ and let $f = (f_1, \ldots, f_r) \in F^r$ be such that $(I_r\partial - A) \cdot f = 0$. Then each component f_i of f is D-finite.* □

Proof We show that there is an operator $P \in K[\partial] \setminus \{0\}$ such that $P \cdot f_i = 0$ for every i. Because of $\partial \cdot f = Af$, we have $\partial M \cdot f = (\sigma(M)A + \delta(M)) \cdot f$ for every $M \in K^{r\times r}$ (Exercise 1). Therefore, by induction, every vector $\partial^k \cdot f$ ($k \in \mathbb{N}$) belongs to the subspace $\{Mf : M \in K^{r\times r}\} \subseteq F^r$, whose dimension is at most $r \times r$. Consequently, the vectors $f, \partial\cdot f, \ldots, \partial^{r^2}\cdot f \in F^r$ are linearly dependent over K, so there are $p_0, \ldots, p_{r^2} \in K$, not all zero, such that $(p_0 + p_1\partial + \cdots + p_{r^2}\partial^{r^2}) \cdot f = 0$. ■

The argument used in the proof is somewhat brutal. It shows not only that each component f_i of a solution of the system is D-finite, but it constructs a single operator P that simultaneously annihilates every component f_i of any solution. This is more than we asked for, and as a result, the implied bound r^2 on the order of P is quite pessimistic. As we shall see later in this section, each component f_i is already annihilated by an operator of order at most r. On the other hand, knowing that there is an operator which annihilates all components of any solution of a coupled system offers us a quick alternative proof for some D-finite closure properties. For example,

consider two D-finite functions $f, g \in F$, and suppose that $P \cdot f = Q \cdot g = 0$ for some $P, Q \in K[\partial] \setminus \{0\}$. Let $C_P \in K^{r\times r}$ and $C_Q \in K^{s\times s}$ be the companion matrices of P and Q, respectively, and consider the system

$$\partial \cdot h = \begin{pmatrix} C_P & \\ & C_Q \end{pmatrix} h.$$

Its solution space contains the vector $(f, g) \in F^2$, and so the operator which annihilates all components of (f, g) must annihilate both f and g. It must therefore annihilate $f + g$, thus showing that $f + g$ is D-finite.

In order to solve a given coupled system, it is not a good idea to work out the argument in the proof of Proposition 4.34. We can get along with shorter equations if we proceed more carefully.

Example 4.35 In order to solve the coupled system

$$\begin{aligned} f'(x) &= \frac{3x^2+4}{x(3x+2)} f(x) - \frac{6(x-2)}{x(3x+2)} g(x) \\ g'(x) &= \frac{3-x}{3x+2} f(x) + \frac{11}{3x+2} g(x), \end{aligned}$$

we can differentiate the first equation to get

$$\begin{aligned} f''(x) = \frac{3x^2+4}{x(3x+2)} f'(x) + \frac{2(3x^2-12x-4)}{x^2(3x+2)^2} f(x) \\ - \frac{6(x-2)}{x(3x+2)} g'(x) + \frac{6(3x^2-12x-4)}{x^2(3x+2)^2} g(x). \end{aligned}$$

Then we can use the second equation to eliminate $g'(x)$. This gives

$$f''(x) = \frac{3x^2+4}{x(3x+2)} f'(x) + \frac{2(3x^3-12x^2+6x-4)}{x^2(3x+2)^2} f(x) - \frac{12(4x^2-5x+2)}{x^2(3x+2)^2} g(x).$$

Finally, use the first equation to eliminate $g(x)$ and obtain

$$f''(x) = \frac{x^2-2}{x(x-2)} f'(x) - \frac{2(x-1)}{x(x-2)} f(x).$$

This equation has a solution space generated by x^2 and $\exp(x)$, and once we choose a solution $f(x) = \alpha x^2 + \beta \exp(x)$, the first equation forces us to set

$$g(x) = \left(\frac{3x^2+4}{x(3x+2)} f(x) - f'(x) \right) \frac{x(3x+2)}{6(x-2)} = -\frac{1}{2}\alpha x^3 - \frac{1}{3}\beta \exp(x).$$

The solution space of the coupled system is therefore generated by $(2x^2, -x^3)$ and $(3\exp(x), -\exp(x))$.

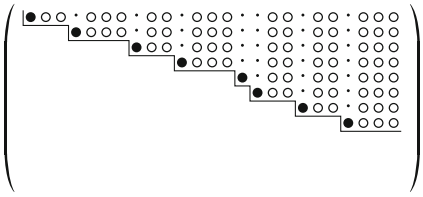
Note that we have found a second order equation for f and another equation that we may regard as a zeroth order inhomogeneous equation for g. The smallest operator that annihilates both f and g has order 3. □

The procedure applied in the example above can be viewed as a kind of Gaussian elimination applied to the matrix $I_r\partial - A \in K[\partial]^{r\times r}$. The usual Gaussian elimination turns a linear system over a field K into an equivalent linear system over K which has a particular structure, the essential property being that if the variables are $x_1, \ldots, x_r$, then for every i there is at most one equation which contains x_i but none of $x_1, \ldots, x_{i-1}$. Although $K[\partial]$ is not a field but a non-commutative ring, we can achieve the same structure for matrices with entries in $K[\partial]$.

Definition 4.36

1. A matrix $A \in K[\partial]^{r\times r}$ is called *(left) unimodular* if there exists a matrix $B \in K[\partial]^{r\times r}$ such that $BA = I_r$. Such a matrix B is then called a *left inverse* of A.
2. A matrix $A = ((a_{i,j}))_{i=1,j=1}^{r,s} \in K[\partial]^{r\times s}$ is said to be in *Hermite normal form* if
   a. For every $j \in \{1, \ldots, s\}$ there is at most one $i \in \{1, \ldots, r\}$ such that $a_{i,j} \neq 0$ and $a_{i,1} = \cdots = a_{i,j-1} = 0$. For this i we have $\mathrm{lc}_\partial(a_{i,j}) = 1$.
   b. For every $(i, j) \in \{1, \ldots, r\} \times \{1, \ldots, s\}$ with $a_{i,j} \neq 0$ and $a_{i,1} = \cdots = a_{i,j-1} = 0$ we have $a_{u,v} = 0$ for all $(u, v) \in \{i+1, \ldots, r\} \times \{1, \ldots, j-1\}$.
   c. For every $(i, j) \in \{1, \ldots, r\} \times \{1, \ldots, s\}$ with $a_{i,j} \neq 0$ and $a_{i,1} = \cdots = a_{i,j-1} = 0$ we have $\mathrm{ord}(a_{1,j}), \ldots, \mathrm{ord}(a_{i-1,j}) < \mathrm{ord}(a_{i,j})$.
3. A matrix $A \in K[\partial]^{r\times s}$ has a matrix $H \in K[\partial]^{r\times s}$ as Hermite normal form if H is a Hermite normal form and there is a left unimodular matrix U such that $UA = H$. □

The technical conditions in the definition of a Hermite normal form express that the matrix should have a staircase shape with zeros below the staircase and with the entries corresponding to a corner of the staircase having the largest order among all entries in their respective column.



Whenever a linear system of operator equations is in Hermite normal form, we can solve it from the bottom up, just like in linear algebra. For every corner position (i, j) in the staircase shape of the matrix, we have an inhomogeneous equation

$$a_{i,j} \cdot f_j = -a_{i,j+1} \cdot f_{j+1} - \cdots - a_{i,s} \cdot f_s$$

whose inhomogeneous part depends on functions $f_{j+1}, \ldots, f_s$ that either have already been determined (if there is a corner position corresponding to their index), or are "arbitrary functions" (if not).

Example 4.37

1. The matrix

$$H = \begin{pmatrix} D-1 & D \\ 0 & D^2 \end{pmatrix}$$

is in Hermite normal form. To solve the system $H \cdot (f_1, f_2)^T = 0$, we first solve $D^2 \cdot f_2 = 0$, which has a solution space generated by 1 and x. For a generic element $f_2 = c_1 + c_2 x$ of the solution space, we next solve the first equation $(D-1) \cdot f_1 = -D \cdot f_2 = c_2$. The space of all triples $(f_1, c_1, c_2) \in C[[x]] \times C^2$ satisfying this equation is generated by $(1, 0, 1)$, $(0, 1, 0)$, and $(\mathrm{e}^x, 0, 0)$. It follows that the solution space of the system $H \cdot (f_1, f_2)^T = 0$ is generated by $(1, x)$, $(0, 1)$, and $(\mathrm{e}^x, 0)$.
2. The matrix

$$H = \begin{pmatrix} D-1 & x & D \\ 0 & D & D^2 \end{pmatrix}$$

is also in Hermite normal form. We want to solve the system $H \cdot (f_1, f_2, f_3)^T = 0$. In this case, the matrix provides no equation for f_3, so we can let f_3 be an arbitrary element of $C[[x]]$. Let $a \in C[[x]]$ be arbitrary. For the choice $f_3 = a$, the component f_2 is determined through the inhomogeneous equation $D \cdot f_2 = -D^2 \cdot a$, whose general solution has the form $f_2 = c - a'$ for an arbitrary constant $c \in C$. Next we consider the equation $(D-1) \cdot f_1 + x f_2 + D \cdot f_3 = 0$. Plugging the general form of f_2 and the choice $f_3 = a$ into the equation leads to $(D-1) \cdot f_1 = -cx - (x+1)a'$. The homogeneous/parametric part $(D-1) \cdot f = -cx$ has a solution space generated by $(\mathrm{e}^x, 0)$ and $(x+1, 1)$ in $C[[x]] \times C$. The solutions of the inhomogeneous equation depend on the choice a and cannot be easily expressed in other terms. We can say however that for every choice a there is a certain $b \in C[[x]]$ such that the solution of the inhomogeneous equation is $f_1 = b + c_1 \mathrm{e}^x + c_2(x+1)$ for certain constants $c_1, c_2 \in C$. The solution set in $C[[x]]^3$ of the entire system $H \cdot (f_1, f_2, f_3)^T = 0$ can be described as

$$\big\{ (b, a', a) + c_1(\mathrm{e}^x, 0, 0) + c_2(x+1, 1, 0) : \\ c_1, c_2 \in C, a \in C[[x]], \text{ and } b \text{ is such that } (D-1) \cdot b = (x+1)a' \big\}.$$

Note that as a C-vector space, this solution set has infinite dimension. □

Faced with a system $A \cdot f = 0$ where $A \in K[\partial]^{r\times s}$ is not in Hermite normal form, we can exploit that for any unimodular matrix $U \in K[\partial]^{r\times r}$ we have $A \cdot f = 0 \iff UA \cdot f = 0$. The idea is thus to successively multiply A by a sequence of unimodular matrices so as to turn A into a Hermite normal form, and then solve the system as illustrated in the example above. This is of course the same general idea as in Gaussian elimination, where the elementary row operations play the role of the unimodular matrices, the only difference being that since $K[\partial]$ is not a field, we cannot simply divide a row by a nonzero matrix entry to produce an element by which all other elements of the column can be eliminated. But we can do division with remainder. If a column contains two nonzero entries, say $a, a' \in K[\partial]$ with $\operatorname{ord}(a) \leq \operatorname{ord}(a')$, we can add the $-\operatorname{rquo}(a', a)$-fold of the row containing a to the row containing a'. This has the effect that a' gets replaced by $\operatorname{rrem}(a', a)$, which must have smaller order than a'. Doing the same computation for all rows in place of the row containing a', we can arrange that a becomes the element of largest order in the column under consideration. Now letting another nonzero entry of the column play the role of a (if there still is one), we can repeat the procedure to ensure that all other entries have strictly lower order. Since orders are natural numbers, we cannot observe infinitely many descents of the maximal order, so after finitely many repetitions we will reach a situation in which there is at most one nonzero entry left in the column. We have then found the first corner of the staircase. We then treat each of the remaining columns in the same way, except that the choice for a is limited to such rows which have no nonzero entries to the left of a. This leads to the following algorithm.

Algorithm 4.38
Input: $A = ((a_{i,j}))_{i=1,j=1}^{r,s} \in K[\partial]^{r\times s}$.
Output: *A Hermite normal form for* A.

Set $k = 1$.
for $j = 1, \ldots, s$ *do*
  while there are $i_1, i_2 \in \{k, \ldots, r\}$ *with* $i_1 \neq i_2$ *and* $a_{i_1,j}, a_{i_2,j} \neq 0$ *do*
    Choose i_1, i_2 *with* $a_{i_1,j}, a_{i_2,j} \neq 0$ *and* $\operatorname{ord}(a_{i_1,j}) \leq \operatorname{ord}(a_{i_2,j})$.
    for $\ell = s, s-1, \ldots, j$ *do*
      $a_{i_2,\ell} = a_{i_2,\ell} - \operatorname{rquo}(a_{i_2,j}, a_{i_1,j})a_{i_1,\ell}$.
  if there is an $i \in \{k, \ldots, r\}$ *with* $a_{i,j} \neq 0$ *then*
    Choose such an i *and swap the* i*th and* k*th row of* A.
    for $\ell = s, s-1, \ldots, j$ *do*
      $a_{k,\ell} = \operatorname{lc}(a_{k,j})^{-1} a_{k,\ell}$.
    for $i = 1, \ldots, k-1$ *and* $\ell = j, \ldots, s$ *do*
      $a_{i,\ell} = a_{i,\ell} - \operatorname{rquo}(a_{i,j}, a_{k,j})a_{k,\ell}$.
    Set $k = k + 1$.
Return A.

Theorem 4.39 *Algorithm 4.38 is correct and terminates.* □

Proof For the termination, the only critical issue is the while loop starting in line 3. Within this loop, $a_{i_2,j}$ gets replaced by $a_{i_2,j} - \operatorname{rquo}(a_{i_2,j}, a_{i_1,j})a_{i_1,j} = \operatorname{rrem}(a_{i_2,j}, a_{i_1,j})$, whose order is strictly smaller than that of $a_{i_1,j}$. No entry of the jth column can get replaced by an element of higher order. The sum of the orders of the entries of the jth column is a natural number which decreases in every iteration. Since this cannot happen infinitely often, the loop must terminate.

For the correctness, note first that there is a unimodular matrix U such that multiplying U from the left to the input matrix produces the output matrix. This is because the matrix is only modified in lines 6, 8, 10 and 12, and the operations performed there correspond to left-multiplications by certain unimodular matrices (Exercise 7). The effect of the entire algorithm corresponds to the product of these matrices.

To show, secondly, that the output is a Hermite normal form, we show by induction on j that at the end of the jth iteration of the loop starting in line 2, the first j columns of A form a Hermite normal form whose first $k-1$ rows are nonzero. For the induction base $j = 1$ there is nothing to show. Suppose the claims are true for some $j-1$ and consider the jth iteration. As the while loop starting in line 3 only affects rows that have only zeros in the first $j-1$ columns, these operations do not affect any entries of these columns. In particular, the Hermite normal form structure of the first $j-1$ columns is preserved. After the while loop, the jth column contains at most one nonzero entry in rows $k, \dots, r$. If it has none, the first j columns form a Hermite normal form with $k-1$ nonzero rows. If there is a nonzero entry, then after executing lines 8–12, the nonzero entry is in row k and monic, and all entries above have lower order. We have thus again a Hermite normal form with k nonzero rows. After updating the counter k in step 13, we have reached the claimed situation. ∎

Example 4.40 Consider the matrix

$$A = \begin{pmatrix} D - \frac{3x^2+4}{x(3x+2)} & \frac{6(x-2)}{x(3x+2)} \\ -\frac{3-x}{3x+2} & D - \frac{11}{3x+2} \end{pmatrix} \in C(x)[D]^{2\times 2},$$

which corresponds to the system already considered in Example 4.35. We compute a Hermite normal form of A using Algorithm 4.38. The first column is cleaned up by adding the $(D - \frac{3x^2+4}{x(3x+2)})\frac{3x+2}{3-x}$-fold of the second row to the first (line 6), exchanging the two rows (line 8), and multiplying the second row by $-\frac{3x+2}{3-x}$ (line 10). The result is

$$\begin{pmatrix} 1 & \frac{3x+2}{x-3}D - \frac{11}{x-3} \\ 0 & \frac{-3x-2}{x-3}D^2 + \frac{(3x+2)(x^2-6)}{x(x-3)^2}D - \frac{3(x-2)(3x+2)}{x(x-3)^2} \end{pmatrix}.$$

After multiplying the second row by $\frac{x-3}{-3x-2}$, we obtain a Hermite normal form of A:

$$\begin{pmatrix} 1 & \frac{3x+2}{x-3}D - \frac{11}{x-3} \\ 0 & D^2 + \frac{6-x^2}{x(x-3)}D + \frac{3(x-2)}{x(x-3)} \end{pmatrix}.$$

□

According to its specification, Algorithm 4.38 computes "a" Hermite normal form for a given matrix in $K[\partial]^{r\times s}$. We show next that for every matrix in $K[\partial]^{r\times s}$ there is at most one Hermite normal form, so that we can meaningfully speak about "the" Hermite normal form of a matrix.

Proposition 4.41 *Let $H_1, H_2 \in K[\partial]^{r\times s}$ be two matrices in Hermite normal form, and suppose that there is a unimodular matrix $U \in K[\partial]^{r\times r}$ such that $UH_1 = H_2$. Then $H_1 = H_2$.* □

Proof We have to show that if there is a U such that $UH_1 = H_2$, we can choose $U = I_r$ as well. We proceed inductively along the structure of a Hermite normal form.

For the base case, note that $H_1 = 0$ if and only if $H_2 = 0$. More generally, the first column of H_1 is zero if and only if the first column of H_2 is. For the induction step, it therefore suffices to consider

$$H_1 = \begin{pmatrix} L_1 & P_1 \\ 0 & Q_1 \end{pmatrix} \quad \text{and} \quad H_2 = \begin{pmatrix} L_2 & P_2 \\ 0 & Q_2 \end{pmatrix},$$

with $L_1, L_2 \in K[\partial] \setminus \{0\}$, $P_1, P_2 \in K[\partial]^{1\times(s-1)}$, and $Q_1, Q_2 \in K[\partial]^{(r-1)\times(s-1)}$. By the induction hypothesis, we have $Q_1 = Q_2$. From $UH_1 = H_2$ we get $u_1L_1 = e_1L_2$, where u_1 is the first column of U and e_1 is the first unit vector. Coefficient comparison implies in succesion: all components of u_1 except for the first are zero, the first component of u_1 has order zero (otherwise U can't be invertible in $K[\partial]^{r\times r}$), $u_1 = e_1$ (using $\operatorname{lc}(L_1) = \operatorname{lc}(L_2)$), and finally $L_1 = L_2$.

We can thus conclude that $U = \begin{pmatrix} 1 & A \\ 0 & I_{r-1} \end{pmatrix}$ for some $A = (a_2, \ldots, a_r) \in K[\partial]^{r-1}$ and it remains to show that we can take $A = 0$. Let $i \in \{2, \ldots, r\}$.

Case 1: The ith row of H_1 is nonzero—say the first nonzero element is $M \in K[\partial]$ and appears in column j. Let h_1, h_2 be the jth columns of H_1, H_2, respectively. By the induction hypothesis, the vectors h_1, h_2 can only differ in their first components $h_{1,1}, h_{2,1}$. More precisely, we have $h_{1,1} + a_iM = h_{2,1}$, and since the structural requirements for a Hermite normal form require $\operatorname{ord}(h_{1,1}), \operatorname{ord}(h_{2,1}) < \operatorname{ord}(M)$, it follows that $a_i = 0$.

Case 2: The ith row of H_1 is zero. In this case, let j be arbitrary let h_1, h_2 be the jth columns of H_1, H_2 with $h_{1,1}, h_{2,1}$ as their (respective) first components. We then have $h_{1,1} + a_i0 = h_{2,1}$, so $h_{1,1}$ and $h_{2,1}$ agree regardless of the choice of a_i and we may take $a_i = 0$. ■

Like for matrices over a field, we can draw some useful conclusions from Proposition 4.41. First of all, a matrix $A \in K[\partial]^{r\times r}$ is unimodular if and only if its Hermite normal form is I_r. Next, a matrix is unimodular if and only if it

can be written as a product of elementary matrices (matrices corresponding to elementary row operations). Finally, since elementary matrices are both left and right unimodular, we find that every left-unimodular matrix is right-unimodular and vice versa, so there is no need to distinguish these notions.

The Hermite normal form has the advantage that we can solve higher order coupled linear systems directly, without having to translate them into first order systems with larger matrices. A disadvantage of the approach is that it is somewhat expensive. There is also an approach for solving linear systems of operator equations without increasing the order or the matrix sizes. Starting from a first order system $(I_r\partial - A) \cdot f = 0$ with $A \in K^{r\times r}$, the idea is to find a basis change that turns A into a companion matrix. If this can be done, the system naturally translates into an equation of order r with coefficients in K, and any solution of this equation can be translated back into a solution of the original system.

It must not be overlooked that the action of ∂ interferes with a basis change. If $P \in K^{r\times r}$ is an invertible matrix and we set $g = Pf$, then

$$\partial \cdot g = \partial \cdot Pf = \sigma(P)(\partial \cdot f) + \delta(P)f = \sigma(P)Af + \delta(P)f = (\sigma(P)A + \delta(P))P^{-1}g$$

so the basis change matrix P transforms the system $(I_r\partial - A) \cdot f = 0$ into the system $(I_r\partial - B) \cdot g = 0$ where $B = (\sigma(P)A + \delta(P))P^{-1} \in K^{r\times r}$. The matrix B is called the *gauge transform* of A with respect to P.

Definition 4.42 Let $K[\partial]$ be an Ore algebra and $P \in K^{r\times r}$ be an invertible matrix. For $A \in K^{r\times r}$, the matrix $P[A] := (\sigma(P)A + \delta(P))P^{-1}$ is called the *gauge transform* of A with respect to P. Two matrices $A, B \in K^{r\times r}$ are *gauge equivalent* if there exists an invertible matrix $P \in K^{r\times r}$ such that $P[A] = B$. □

Example 4.43

1. Continuing the previous example, let

$$A = \begin{pmatrix} \frac{3x^2+4}{x(3x+2)} & -\frac{6(x-2)}{x(3x+2)} \\ \frac{3-x}{3x+2} & \frac{11}{3x+2} \end{pmatrix} \in C(x)^{2\times 2}$$

and consider the system $(I_2D - A) \cdot f = 0$. With

$$P = \begin{pmatrix} 1 & 0 \\ \frac{3x^2+4}{x(3x+2)} & -\frac{6(x-2)}{x(3x+2)} \end{pmatrix} \in C(x)^{2\times 2}$$

we have

$$P[A] = \begin{pmatrix} 0 & 1 \\ -\frac{3(x-2)}{x(x-3)} & -\frac{6-x^2}{x(x-3)} \end{pmatrix}.$$

The matrix $P[A]$ is the companion matrix of the operator $D^2 + \frac{6-x^2}{x(x-3)}D + \frac{3(x-2)}{x(x-3)} \in C(x)[D]$. The solution x^2 of this operator gives rise to the solution $(x^2, 2x)$ of the coupled system $(I_2 D - P[A]) \cdot f = 0$, which in turn gives rise to the solution $P^{-1}(x^2, 2x) = (x^2, x^3/2)$ of the original system $(I_2 D - A) \cdot f = 0$. Likewise, the solution $\exp(x)$ of the operator translates into the solution $P^{-1}(\exp(x), \exp(x)) = (\exp(x), -\exp(x)/3)$.

2. For $L = D^2 - x \in C(x)[D]$, the module $C(x)[D]/\langle L\rangle$ is a $C(x)$-vector space of dimension 2. The sets $B_1 = \{1, D\}$ and $B_2 = \{x + D, x - D\}$ are bases of this vector space. For transforming a B_1-representation of some element of $C(x)[D]/\langle L\rangle$ into a B_2-representation of the same element, we do *not* need a gauge transform. Instead, this is a matter of the usual matrix-vector multiplication from linear algebra. □

Given an arbitrary matrix $A \in K^{r\times r}$, our goal is to find an invertible matrix $P \in K^{r\times r}$ such that $P[A]$ has the form of a companion matrix. Suppose $P \in K^{r\times r}$ is such a matrix, i.e., such that $P[A] = (\sigma(P)A + \delta(P))P^{-1} = C_L$, where

$$C_L = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ * & \cdots & \cdots & \cdots & * \end{pmatrix} \in K^{r\times r}$$

is the companion matrix of some operator $L \in K[\partial]$. We then have $\sigma(P)A+\delta(P) = C_L P$, and by the shape of a companion matrix, the ith row of $C_L P$ is equal to the $(i + 1)$st row of P, for $i = 1, \ldots, r - 1$. At the same time, if we know the ith row of P, we can compute from it the ith row of $\sigma(P)A + \delta(P)$, so if we have $\sigma(P)A + \delta(P) = C_L P$, we can compute the $(i + 1)$st row of P from the ith row of P, for every $i = 1, \ldots, r - 1$. In other words, it suffices to determine the first row of a suitable transformation matrix P. The remaining rows of P are uniquely determined.

Conversely, let $p \in K^r$ be any vector (viewed as a row vector) and consider the matrix $P \in K^{r\times r}$ whose rows are $p_1, \ldots, p_r \in K^r$ defined by $p_1 = p$ and $p_{i+1} = \sigma(p_i)A + \delta(p_i)$ $(i = 1, \ldots, r - 1)$. We then have

$$\begin{pmatrix} \sigma(p_1) \\ \vdots \\ \vdots \\ \sigma(p_{r-1}) \end{pmatrix} A + \begin{pmatrix} \delta(p_1) \\ \vdots \\ \vdots \\ \delta(p_{r-1}) \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}}_{\in K^{(r-1)\times r}} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ \vdots \\ p_r \end{pmatrix},$$

and it remains to check whether there is some vector $u \in K^r$ which we can use as an additional rth row in the first matrix on the right so that the left hand side gains $\sigma(p_r)A + \delta(p_r)$ as an rth row. The question amounts to an inhomogeneous linear system for the unknown vector u, and since we are only interested in situations where P is invertible, we can be sure that a unique vector u exists.

Whether P is invertible depends on the choice of its first row p. A vector $p \in K^r$ is called *cyclic* (with respect to A, σ, δ) if the matrix $P \in K^{r\times r}$ constructed as described above is invertible. Equivalently, p is cyclic if and only if the vectors $p_1, \ldots, p_r \in K^r$ form a basis of K^r. We could now proceed to argue that almost all vectors are cyclic and propose a randomized algorithm that picks candidates at random until a cyclic vector is encountered. Some implementations proceed like this. In fact, it is not really so dramatic if we accidentally encounter a non-cyclic vector. If p is not cyclic, i.e., if $p_1, \ldots, p_r \in K^r$ do not form a basis of K^r, then there is some $i < r$ such that $p_1, \ldots, p_i$ are linearly independent but $p_{i+1} = \sigma(p_i)A + \delta(p_i)$ is a K-linear combination of $p_1, \ldots, p_i$. We can adjust the definition of the vectors $p_1, p_2, \ldots$ by setting $p_{i+1} = \sigma(p_i)A + \delta(p_i)$ only if this vector does not belong to the K-linear subspace of K^r generated by $p_1, \ldots, p_i$, and otherwise setting p_{i+1} to an arbitrary vector that does not belong to the subspace generated by $p_1, \ldots, p_i$. With this definition, it is clear that the resulting vectors $p_1, \ldots, p_r$ will be a basis of K^r, and if $P \in K^{r\times r}$ is the matrix which has $p_1, \ldots, p_r$ as rows, then we have $\sigma(P)A + \delta(P) = A'P$ for some matrix $A' \in K^{r\times r}$ which has the form

$$A' = \begin{pmatrix} \boxed{\begin{matrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ * & \cdots & \cdots & * \end{matrix}}\ \} r_1 \times r_1 & & \\ & \boxed{\begin{matrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ * & \cdots & \cdots & * \end{matrix}}\ \} r_2 \times r_2 & \\ & & \boxed{\begin{matrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ * & \cdots & \cdots & * \end{matrix}} \\ & & & \ddots \end{pmatrix}.$$

This is in general not a companion matrix, but it is close enough. If there are m companion-like blocks in A' of respective sizes $r_1, \ldots, r_m$, then the system $(I_r\partial - A') \cdot f = 0$ translates into a system

$$\begin{aligned} L_1 \cdot f_1 &= 0 \\ L_2 \cdot f_2 &= M_{2,1} \cdot f_1 \\ &\vdots \end{aligned}$$

$$L_m \cdot f_m = M_{m,1} \cdot f_1 + \cdots + M_{m,m-1} \cdot f_{m-1}$$

in which $L_1, \ldots, L_m \in K[\partial]$ are known operators of respective orders $r_1, \ldots, r_m$, the $M_{i,j} \in K[\partial]$ are known operators of order $< r_j$, and $f_1, \ldots, f_m$ are unknown scalar functions. The system is uncoupled in the sense that we can solve it if we know how to solve (inhomogeneous) scalar equations.

If we choose the kth unit vector as the first p, then the resulting operator L_1 corresponding to the first block will be an annihilating operator for the kth component of any solution vector $f \in K^r$ of the original system $(I_r\partial - A) \cdot f = 0$. Since $\operatorname{ord}(L_1) \leq r$, we see that each component of a solution vector of such a system is D-finite of order at most r. This refines the bound obtained in Proposition 4.34.

The uncoupling algorithm described above can be summarized as follows:

Algorithm 4.44
Input: $A \in K^{r\times r}$, *an endomorphism* $\sigma\colon K \to K$, *and a* σ*-derivation* $\delta\colon K \to K$.
Output: *An invertible matrix* $P \in K^{r\times r}$ *such that* $P[A]$ *has the shape described above.*

1 *Set* $P = 0 \in K^{r\times r}$ *and* $k = 0$.
2 *while* $k < r$ *do*
3 *Choose a vector* $p \in K^{r\times r} \setminus \{0\}$ *and set the* $(k+1)$*st row of* P *to* p.
4 *for* $i = 2, \ldots, r-k$ *do*
5 *Set* $p = \sigma(p)A + \delta(p)$ *and set the* $(k+i)$*th row of* P *to* p.
6 *Set* k *to be the largest number such that the first* k *rows of* P *are linearly independent over* K.
7 *Return* P.

Theorem 4.45

1. *Algorithm 4.44 is correct.*
2. *Suppose that* $K = C(x)$ *and* σ, δ *map polynomials to polynomials with* $\deg \sigma(p), \deg \delta(p) < \deg p$ *for all* $p \in C[x]$, *and suppose that the application of* σ *and* δ *to a polynomial of degree* d *costs no more than* $\mathrm{O}^{\sim}(d)$ *operations in* C. *For this setting, Algorithm 4.44 can be implemented in such a way that whenever it is applied to any matrix* $A \in C[x]^{r\times r}$ *with entries of degree at most* d, *it performs no more than* $\mathrm{O}^{\sim}(r^{\omega+2}d)$ *operations in* C. □

Proof

1. That P is invertible follows from the choice of k in line 6 and the termination condition in line 2. That $P[A]$ has the required form follows from the discussion above. This implies the correctness of the algorithm.
2. We need to be more specific about the implementation of lines 3 and 6. If in line 3 we always choose a vector with entries in C, then P will always be a matrix with entries of degree at most rd.
   For line 6, we can apply a bisection search to find k such that the top k rows of P form a matrix of rank k. This takes at most $\log(k)$ rank computations,

each of which can be done with at most $O^{\sim}(r^{\omega}rd)$ operations in C according to Theorem 1.28. Since $\log(k) \leq \log(r)$, identifying k also costs $O^{\sim}(r^{\omega}rd)$ operations. If k increases in each iteration of the loop, the total cost contributed by line 6 amounts to $O^{\sim}(r^{\omega+2}d)$ operations.
In order to ensure that k increases in each loop iteration, we should choose in line 3 a vector p which does not belong to the vector space generated by the first k rows of P. One way of doing so is to maintain a pool of candidates, which is initially set to the set of all the unit vectors $e_1, \ldots, e_r$. In line 3, we select an element from the pool and test whether it is linearly independent with the first k rows of P. This costs $O^{\sim}(r^{\omega}rd)$ operations. If it is linearly dependent, it will always be, so we can safely remove it from the pool and try another element. Once we find an element that is linearly independent, we take it as p and remove it from the pool. Since the r vectors initially in the pool form a basis of K^r, we will never run out of candidates. Moreover, we will altogether at most r times check whether a pool element is suitable, so the total number of operations spent in line 3 can be limited to $O^{\sim}(r^{\omega+2}d)$.
With the assumptions on σ and δ, each execution of line 5 costs $O^{\sim}(r^2d)$ operations, so each execution of the loop in lines 4 and 5 amounts to $O^{\sim}(r^3d)$ operations, and the total cost to $O^{\sim}(r^4d)$. Since $\omega \geq 2$, this is bounded by $O^{\sim}(r^{\omega+2}d)$ operations. ∎

Algorithm 4.44 pays a price for being a deterministic algorithm. If we are willing to turn it into a randomized algorithm, we could simply let it choose a random element of K^r as p, build a candidate transformation matrix P from it, and check whether it is invertible. With high probability, this will be the case. If it is not the case, we can either return "failed" or try again, depending on whether we prefer a Monte Carlo or a Las Vegas style randomized algorithm. Either way, the expected runtime drops to $O^{\sim}(r^{\omega+1}d)$ for the setting described in part 2 of Theorem 4.45.

Being deterministic, Algorithm 4.44 has the feature that we can also choose simple vectors as p in line 3, which might not qualify as honest random elements of K^r. This way, we can keep the degrees of the entries in P low.

Coupled systems of functional equations arise naturally in a number of contexts. As an example, let $K[\partial]$ be an operator and $L_1, L_2 \in K[\partial]$ be two elements of some order r, and consider the corresponding quotient modules $M_1 = K[\partial]/\langle L_1 \rangle$ and $M_2 = K[\partial]/\langle L_2 \rangle$. Given L_1, L_2, how can we decide whether M_1 and M_2 are isomorphic as $K[\partial]$-modules? Since the module M_1 is generated by $[1]$, a module homomorphism $h\colon M_1 \to M_2$ is uniquely determined by an operator $U \in K[\partial]$ such that $h([1]) = [U] \in M_2$. For more clarity, instead of the generic equivalence class notation $[P]$, we will write $[P]_{L_i}$ for the class of P modulo L_i, so that $[P]_{L_i}$ is more easily recognized as an element of M_i $(i = 1, 2)$. For any other operator $P \in K[\partial]$ we have

$$[PU]_{L_2} = P \cdot [U]_{L_2} = P \cdot h([1]_{L_1}) = h(P \cdot [1]_{L_1}) = h([P]_{L_1}).$$

In order for h to be well-defined, we must ensure that the zero of M_1 is mapped to the zero of M_2, i.e., that $h([0]_{L_1}) = h([L_1]_{L_1}) = [L_1U]_{L_2} = [0]_{L_2}$, i.e., that $L_1U = VL_2$ for some $V \in K[\partial]$. We can search for U and V simultaneously by making an ansatz $U = u_0 + \cdots + u_{r-1}\partial^{r-1}$, $V = v_0 + \cdots + v_{r-1}\partial^{r-1}$ and equate the coefficients of $L_1U - VL_2$ to zero. This looks similar to the computation of least common left multiples, but observe that U is now to the right of L_1 rather than to the left. By commuting the powers of ∂ appearing in L_1 with the undetermined coefficients of U, we introduce derivations of these, so that the coefficient comparison does not simply result in a linear system over K but in a system of functional equations.

The solutions of the functional equations give rise to the choices of U that lead to well-defined homomorphisms from M_1 to M_2. In order to decide the isomorphism question, it remains to check whether any of these homomorphisms is surjective. For a specific choice U, this is the case if and only if the elements $[\partial^i U]_{L_2}$ $(i = 0, \ldots, r-1)$ are linearly independent over K, which is easy to check with linear algebra. The set of all possible choices U forms a finite-dimensional C-vector space, generated by $U_1, \ldots, U_m$. In order to check whether this space contains an element that corresponds to an isomorphism, consider a linear combination $U = c_1U_1 + \cdots + c_mU_m$ with undetermined coefficients $c_1, \ldots, c_m$ and use linear algebra to find out whether the elements $[\partial^i U]_{L_2}$ $(i = 1, \ldots, r-1)$ are linearly dependent for every choice of $c_1, \ldots, c_m$. This is the case if and only if there is no isomorphism.

Example 4.46

1. Let $L_1 = D^2 - x$ and $L_2 = (1-x)D^2 + D + (x^2 - x - 1)$, and consider the modules $M_1 = C(x)[D]/\langle L_1 \rangle$ and $M_2 = C(x)[D]/\langle L_2 \rangle$. For deciding whether $M_1 \cong M_2$ as $C(x)[D]$-modules, we make an ansatz

$$(D^2 - x)(u_0 + u_1 D) - (v_0 + v_1 D)((1-x)D^2 + D + (x^2 - x - 1)) = 0$$

with undetermined coefficients $u_0, u_1, v_0, v_1 \in C(x)$. Expanding and collecting terms gives

$$\begin{aligned}
&\big(u_0'' - xu_0 + (-x^2 + x + 1)v_0 + (1 - 2x)v_1\big) \\
&+ \big(2u_0' + u_1'' - xu_1 - v_0 + (-x^2 + x + 1)v_1\big)D \\
&+ \big(u_0 + 2u_1' + (x-1)v_0\big)D^2 + \big(u_1 + (x-1)v_1\big)D^3 = 0.
\end{aligned}$$

By equating the coefficients of D^2 and D^3 to zero, we can solve for v_0, v_1 in terms of u_0, u_1, u_1', and plugging these expressions into the coefficients of D^0 and D^1 gives a coupled system of differential equations for u_0 and u_1:

$$\begin{aligned}
(x-1)u_0'' + 2(x^2 - x - 1)u_1' - u_0 + (2x-1)u_1 &= 0, \\
(x-1)u_1'' + 2(x-1)u_0' + u_0 + 2u_1' - u_1 &= 0.
\end{aligned}$$

With the methods described in this section, we can determine the solution space of this system. It turns out to be generated by $(u_0, u_1) = (\frac{1}{1-x}, \frac{1}{1-x}) \in C(x)^2$. It remains to check whether for $U = \frac{1}{1-x}(1+D)$ the operators $\operatorname{rrem}(U, L_2)$ and $\operatorname{rrem}(DU, L_2)$ are linearly independent over $C(x)$. Since they are, it follows that M_1 and M_2 are isomorphic.

2. Now let $L_1 = D^2 - (x+1)D + x$ and $L_2 = (2-x^2)D^2 + 2xD + (x^2-6)$, and consider the modules $M_1 = C(x)[D]/\langle L_1\rangle$ and $M_2 = C(x)[D]/\langle L_2\rangle$. In this case, the ansatz $L_1(u_0 + u_1D) - (v_0 + v_1D)L_2 = 0$ leads to a system of equations which, after eliminating v_0 and v_1, reads

$$\begin{aligned}
&(x^2-2)u_0'' - (x+1)(x^2-2)u_0' + (x^3+x^2-2x-6)u_0 \\
&\qquad + (2x^2-6)u_1' - (x+3)(x^2-2x-2)u_1 = 0, \\
&(2x^2-2)u_0' + (-x^3-x^2+4x+2)u_0 \\
&\qquad + (x^2-2)u_1'' + (-x^3-x^2+6x+2)u_1' + (x^3-x^2-4x-4)u_1 = 0.
\end{aligned}$$

Its solution space is generated by $(u_0, u_1) = (\frac{x-2}{x^2-2}, \frac{x-1}{x^2-2})$, but for $U = \frac{x-2}{x^2-2} + \frac{x-1}{x^2-2}D$ we have $\operatorname{rrem}(DU, L_2) = U$, which is obviously linearly dependent with U. It follows that although there is a nontrivial homomorphism from M_1 to M_2, the modules are not isomorphic.

3. For $L_1 = D^2 - x$ and $L_2 = D^2 + x$ the resulting coupled system has only the solution $(u_0, u_1) = (0, 0)$, so in this case, there is no nontrivial homomorphism from $C(x)[D]/\langle L_1\rangle$ to $C(x)[D]/\langle L_2\rangle$. □

In order to solve a coupled system, we uncouple it so that algorithms from earlier chapters become applicable. We have seen that one way of uncoupling is to apply a suitable gauge transformation to the system. Gauge transformations are not only useful for uncoupling, but they can also be used to study other aspects of the system at hand. For example, gauge transformations are used for extending the definition of removable singularities to systems, and for detecting them. For simplicity, let us restrict to the shift case. In this case, an element of $C/\mathbb{Z}$ is called a *singularity* of a system $(I_r S - A) \cdot f = 0$ with $A \in C(x)^{r\times r}$ if it contains a pole of an entry of A. A singularity is called *removable* if there is a polynomial matrix $P \in C[x]^{r\times r}$ with nonzero determinant (so that it is invertible as element of $C(x)^{r\times r}$) such that $P[A]$ does not have this singularity.

Example 4.47 Let

$$A = \begin{pmatrix} \frac{3x^3-4x^2-2x-2}{x(x+1)} & -\frac{x^3-7x^2-2x-2}{2x(x+1)} \\ \frac{2(3x^3-4x^2-x-1)}{x(x+1)} & -\frac{2x^3-6x^2-x-1}{x(x+1)} \end{pmatrix} \in C(x)^{2\times 2}$$

and consider the system $(I_2 S - A) \cdot f = 0$. Its only singularity is the class $\mathbb{Z} \in C/\mathbb{Z}$. To see whether it is removable, we first try to eliminate the factors $x+1$ and then the

factors x from the denominators. Useful gauge transforms to this end are constant matrices and diagonal matrices with polynomial entries. Using constant matrices, we can try to remove poles by applying suitable linear combinations to the rows and columns of the matrix at hand. For example, since

$$[(x+1)^{-1}]A = \begin{pmatrix} 7 & -4 \\ 14 & -8 \end{pmatrix},$$

we can make some progress by adding the (-2)-fold of the first row to the second. Note that this has the side effect that the 2-fold of second column gets added to the first, but this will not spoil the desired elimination effect.

$$P_1 = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \quad \Rightarrow \quad A_1 := \sigma(P_1) A P_1^{-1} = \begin{pmatrix} \frac{x(2x+3)}{x+1} & -\frac{x^3-7x^2-2x-2}{2x(x+1)} \\ -2x & -\frac{x^2+1}{x} \end{pmatrix}.$$

We have successfully removed the factor $x+1$ from the denominators of the second row. To remove them also from the first row, we can simply multiply this row by $x+1$. This has the side effect that the first column will be divided by x, but we are lucky that x appears in all numerators of the first column, so that no new denominators get introduced. If this were not the case, we would nevertheless proceed in the same way, because we will also have to deal with the denominators x in the second column.

$$P_2 = \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} \quad \Rightarrow \quad A_2 := \sigma(P_2) A_1 P_2^{-1} = \begin{pmatrix} 2x+3 & -\frac{x^3-7x^2-2x-2}{2x} \\ -2 & -\frac{x^2+1}{x} \end{pmatrix}.$$

To get rid of the remaining denominators, considering

$$[x^{-1}]A_2 = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}$$

suggests adding the first column to the second. This gives

$$P_3 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \Rightarrow \quad A_3 := \sigma(P_3) A_2 P_3^{-1} = \begin{pmatrix} \frac{(x-1)(x^2-2x+2)}{2x} & -\frac{x^3-7x^2-2x-2}{2x} \\ \frac{1}{2}(x-1)x & \frac{1}{2}(-x^2+5x+2) \end{pmatrix},$$

from which we clear denominators by multiplying the first row with x:

$$P_4 = \begin{pmatrix} x-1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Rightarrow \quad A_4 := \sigma(P_4) A_3 P_4^{-1} = \begin{pmatrix} \frac{1}{2}(x^2-2x+2) & \frac{1}{2}(-x^3+7x^2+2x+2) \\ \frac{x}{2} & \frac{1}{2}(-x^2+5x+2) \end{pmatrix}.$$

Since we have reached a matrix with polynomial entries, the singularity $\mathbb{Z}$ of A is removable. The matrix $P = P_4P_3P_2P_1 \in C[x]^{2\times 2}$ is a gauge transform which removes the singularity from A. □

Exercises

1. Let $K[\partial]$ be an Ore algebra. Show that $K^{r\times r}[\partial]$ is an Ore algebra if σ and δ are defined entry-wise by the σ and δ of $K[\partial]$.

2. Let $K[\partial]$ be an Ore algebra and $A \in K^{r\times r}$ be invertible. Show: **a.** $\sigma(A^{-1}) = \sigma(A)^{-1}$; **b.** $\delta(A^{-1}) = -\sigma(A^{-1})\delta(A)A^{-1}$.

3. Consider the Ore algebra $C[\partial]$ (i.e., $\sigma = \mathrm{id}$ and $\delta = 0$). Show that for every $A \in C^{r\times r}$ there is an operator $L \in C[\partial]$ of order at most r such that every component of a solution $f \in F^r$ of the system $(I_r\partial - A)\cdot f = 0$ is annihilated by L. This refines the bound of Proposition 4.34 for this particular situation.

4. Show that for every $r \in \mathbb{N}$ there exists a matrix $A \in C(x)^{r\times r}$ such that every operator $L \in C(x)[D]\setminus\{0\}$ which annihilates each component of a solution $f \in F^r$ of the system $(I_rD - A)\cdot f = 0$ has order at least r^2.

5⋆. In the shift case, reprove that D-finiteness is preserved under multiplication using Proposition 4.34.

6. Show that $\begin{pmatrix} S+1 & S^2 - x(x+2)S + x \\ 1 & S - x^2 \end{pmatrix} \in C(x)[S]^{2\times 2}$ is unimodular.

7. **a.** Show that every invertible matrix $A \in K^{r\times r}$ is unimodular as an element of $K[\partial]^{r\times r}$.

b. Let $r \in \mathbb{N}$, $u, v \in \{1,\dots,r\}$, $u \neq v$, and let $L \in K[\partial]$. Let $A = ((a_{i,j}))_{i,j=1}^{r} \in K[\partial]^{r\times r}$ be defined by $a_{u,v} = L$ and $a_{i,j} = \delta_{i,j}$ when $(i,j) \neq (u,v)$. Show that A is unimodular.

8. Find all solutions in $C(x)^3$ of the following systems:

a. $(I_3S - \frac{1}{2x(1+x)}\begin{pmatrix} x(2x+1) & -4x(x+1) & -1 \\ 0 & -2x(x+1) & 0 \\ -x(x+1) & 4x(x+1)^2 & (x+1)(2x+1) \end{pmatrix})\cdot f = 0$

b. $(I_3D - \frac{1}{2x^2}\begin{pmatrix} -x & x^2 & -1 \\ 0 & 2x^2 & 0 \\ -x^2 & -x^3 & x \end{pmatrix})\cdot f = 0$

9. Construct a matrix $A \in C(x)^{3\times 3}$ so that the solution space of the system $(I_3D - A)\cdot f = 0$ in $C(x)^3$ is generated by $\begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}, \begin{pmatrix} 1 \\ x+1 \\ (x+1)^2 \end{pmatrix}, \begin{pmatrix} 1 \\ x+2 \\ (x+2)^2 \end{pmatrix}$.

10. Let $A \in K[\partial]^{r\times 1} \setminus \{0\}$ and let H be the Hermite normal form of A. Show that $H = (G, 0, \ldots, 0)^T$ where $G \in K[\partial]$ is the greatest common right divisor of the entries of A.

11. In this section, we have only discussed homogeneous systems $(I_r\partial - A)\cdot f = 0$. We want to adapt the methods to solve parameterized inhomogeneous systems $(I_r\partial - A)\cdot f = c_1 g_1 + \cdots + c_m g_m$ where $A \in K^{r\times r}$ and $g_1, \ldots, g_m \in K^r$ are given and $f \in K^r$ and $c_1, \ldots, c_m \in C$ are unknown. How can we do this **a.** using the Hermite normal form; **b.** using Algorithm 4.44?

12⋆. Show that gauge equivalence is an equivalence relation.

13. Design an algorithm for the differential case with $K = C(x)$ that decides whether two given matrices A, B are gauge equivalent.

14⋆⋆. Prove or disprove:

a. Any two gauge equivalent systems have the same solution space.

b. Any two companion matrices which are gauge equivalent are in fact equal.

c. For $A = I_r$ there is no cyclic vector when $r \geq 2$.

15. Show that whenever $A \in K^{r\times r}$ is a companion matrix, e_1 is a cyclic vector. Find a matrix $A \in K^{r\times r}$ which is not a companion matrix but for which e_1 is nevertheless a cyclic vector.

16. Let $L \in K[\partial]$ with $r = \operatorname{ord}(L) > 0$ and let $C_L \in K^{r\times r}$ be the companion matrix of L. Show that the definition $\partial \cdot p := \sigma(p)C_L + \delta(p)$ turns the K-vector space K^r into a $K[\partial]$-module which is isomorphic to $K[\partial]/\langle L\rangle$.

17⋆. Let $K[\partial]$ be an Ore algebra and F be a $K[\partial]$-module such that for every $L \in K[\partial]$ of order r the solution space $V(L)$ has dimension at most r (as vector space over C). Show that for every $A \in K^{r\times r}$ the solution space of the system $(I_r\partial - A)\cdot f = 0$ in F^r has dimension at most r (as vector space over C).

18⋆. Show that for every $A \in C[[x]]^{r\times r}$ the system $(I_r D - A)\cdot f = 0$ has r linearly independent solutions in $C[[x]]^r$.

19⋆. Let $A = \begin{pmatrix} 0 & 1/x \\ 0 & 0 \end{pmatrix} \in C(x)^{2\times 2}$. Show that 0 is not a removable singularity in the differential case.

Hint: Consider the solutions of the system $(I_2 D - A)\cdot f = 0$ and use the results of the previous two exercises.

20⋆⋆. Gauge transformations can be defined not only for matrices, but also for operators. For $P, A \in K[\partial]$ with $\operatorname{lc}(P) = 1$, the operator $P[A] := \operatorname{rquo}(\operatorname{lclm}(P, A), P) \in K[\partial]$ is called the *gauge transform* of A with respect to P. Suppose that $K[\partial]$ acts on a module F such that $\dim V(L) = \operatorname{ord}(L)$ for every $L \in K[\partial]$. Show the following properties of the gauge transform:

a. $V(P[A]) = P \cdot V(A)$ (in other words, P acts as a C-vector space isomorphism from $V(A)$ to $V(P[A])$).

b. $P[\operatorname{lclm}(A, B)] = \operatorname{lclm}(P[A], P[B])$.

c. $\operatorname{gcrd}(P, A) = 1 \iff \operatorname{ord}(A) = \operatorname{ord}(P[A])$.

21⋆⋆. The Bessel function $J_\nu(x)$ satisfies the differential equation $x^2 J_\nu''(x) + x J_\nu'(x) + (x^2 - \nu^2) J_\nu(x) = 0$. Solve the differential equation

$$x(2x-1)(10x+9)f''(x)+2(50x^2+39x-18)f'(x)+(20x^3+8x^2-3x+99)f(x) = 0$$

in terms of Bessel functions. More precisely, find $u, v \in C(x)$ such that $f(x) = u(x)J_2(x) + v(x)J_2'(x)$ is a solution.

22. Design an integration algorithm for algebraic functions. More precisely, for a given algebraic extension $K = C(x)[y]/\langle m \rangle$ of $C(x)$ and a given element $f \in K$, the algorithm shall decide whether there exists a $g \in K$ such that $g' = f$ (i.e., $\int f = g$).

Hint: Recall that K is a finite-dimensional $C(x)$-vector space.

23. A sequence $(p_n)_{n=0}^\infty$ is called a *quasi-polynomial* if there is a root of unity $\omega \in C$, say of order $k \in \mathbb{N}$, and polynomials $p_0, \ldots, p_{k-1} \in C[x]$ such that $p_n = p_0(n)+\omega^n p_1(n)+\cdots+\omega^{(k-1)n} p_{k-1}(n)$ for all $n \in \mathbb{N}$. Let $(p_n^{(0)})_{n=0}^\infty, \ldots, (p_n^{(r)})_{n=0}^\infty$ be quasi-polynomials and suppose that the sequence $(a_n)_{n=0}^\infty$ satisfies the recurrence

$$p_n^{(0)} a_n + p_n^{(1)} a_{n+1} + \cdots + p_n^{(r)} a_{n+r} = 0$$

for all $n \in \mathbb{N}$. Show that $(a_n)_{n=0}^\infty$ is D-finite.

References

Algorithm 4.38 is perhaps the most straightforward algorithm for computing the Hermite normal form, but it is certainly not the most efficient one. A less straightforward but more efficient algorithm was proposed by Giesbrecht and Kim [215]. Their algorithm only requires a polynomial number of operations in the constant field C.

Churchill and Kovacic [152] prove in a fairly general setting that cyclic vectors always exist, and that in fact almost every vector is cyclic. Their paper contains references to several earlier proofs. While the cyclic vector method is the oldest way to uncouple systems, it was long considered not satisfactory, so alternative methods were developed, for example by Barkatou [41], by Zürcher [477] and by Abramov and Zima [20]. Bostan, Chyzak and de Panafieu [92] somewhat rehabilitated the cyclic vector approach by a careful complexity analysis.

Our main motivation for uncoupling algorithms is that we want to apply the algorithms from Chaps. 2 and 3 to find the solution of systems. It is also possible to compute such solutions directly, without first transforming the given system into a scalar equation. Such algorithms were developed by Barkatou [42] for the differential case and by Abramov and Barkatou [8] for the recurrence case. These algorithms have been implemented as Maple package ISOLDE by Barkatou and

Pflügel. The analysis of removable singularities sketched at the end of the section is another example for a problem that can be solved without uncoupling. Complete desingularization algorithms for systems were given by Barkatou and Maddah [45] for the differential case, and by Barkatou and Jaroschek [43, 44] for the shift case.

4.4 Factorization

If $f \in F$ is annihilated by some operator $L \in K[\partial]$, then it is also annihilated by every left multiple ML of L, because $(ML)\cdot f = M\cdot(L\cdot f) = M\cdot 0 = 0$. Conversely, if we are interested in "simple" solutions of a given operator $L \in K[\partial]$, we could try to write the operator as a product $L = L_1L_2$ of two operators $L_1, L_2 \in K[\partial]$, because every solution of L_2 will also be a solution of L. We have already done so in Sects. 2.6 and 3.6, when we searched for hypergeometric or hyperexponential solutions of a given equation. We have seen that such solutions correspond to right factors of order 1. If there are no hypergeometric or hyperexponential solutions, i.e., no right factors of order 1, the next natural question is whether there are right factors of higher order. In the present section we discuss how this question can be answered.

Definition 4.48 An operator $L \in K[\partial] \setminus K$ is called *irreducible* if for any $P, Q \in K[\partial]$ with $L = PQ$ we have $\operatorname{ord}(P) = 0$ or $\operatorname{ord}(Q) = 0$. If it is not irreducible, it is called *reducible.* □

As in the case of commutative polynomial rings, every operator $L \in K[\partial] \setminus K$ can be written as a product of finitely many irreducible factors. However, unlike in the commutative case, the factorization is in general not unique. In fact, there may be infinitely many different factorizations, and it is not hard to see why. Consider for example the operator $L = D^2 \in C(x)[D]$. Every polynomial $\alpha + x \in C[x]$ is annihilated by D^2, and since $\alpha + x$ is also annihilated by $D - \frac{1}{\alpha+x}$, the operator $\operatorname{gcrd}(D^2, D - \frac{1}{\alpha+x}) = D - \frac{1}{\alpha+x}$ must be a nontrivial right factor of D^2, for every choice $\alpha \in C$. Another factorization is of course $L = DD$.

Although this example seems to indicate the opposite, it turns out that the factorization of an operator is essentially unique. In order to see in which sense, we will view the factorization of operators as structural properties of modules. For a given operator $L \in K[\partial]$, consider the module $K[\partial]/\langle L\rangle$. If L admits a nontrivial factorization $L = AB$, then the equivalence class $[B]$ generates a nontrivial submodule of $K[\partial]/\langle L\rangle$: it is the K-vector space generated by $[B], [\partial B], \ldots, [\partial^{\operatorname{ord}(A)-1} B]$. Conversely, suppose that $K[\partial]/\langle L\rangle$ has a nontrivial submodule (i.e., a submodule other than $\{0\}$ and $K[\partial]/\langle L\rangle$), and let $B \in K[\partial]$ be such that $[B]$ is one of its elements. Then $[B], [\partial B], \ldots$ generate a K-subspace of $K[\partial]/\langle L\rangle$ of dimension less than $\operatorname{ord}(L)$, say of dimension s. This means that the elements $[B], \ldots, [\partial^s B]$ of $K[\partial]/\langle L\rangle$ are linearly dependent over K, so there is an operator $A \in K[\partial]$ of order s with $A\cdot[B] = 0$. In other words, $\operatorname{rrem}(AB, L) = 0$, or $AB = QL$ for yet another operator $Q \in K[\partial]$. Because of $\operatorname{ord}(A) = s < \operatorname{ord}(L)$,

we have $\operatorname{ord}\operatorname{lclm}(B, L) < \operatorname{ord}(B) + \operatorname{ord}(L)$, which by Exercise 13 of Sect. 4.2 implies that $\operatorname{ord}\gcrd(B, L) > 0$, so L has a nontrivial right factor.

In summary, we have shown that $L \in K[\partial] \setminus K$ is irreducible if and only if the module $K[\partial]/\langle L\rangle$ is *simple,* meaning its only submodules are $\{0\}$ and $K[\partial]/\langle L\rangle$. Finding a right factor of $L \in K[\partial] \setminus K$ is therefore equivalent to finding a nontrivial submodule of $K[\partial]/\langle L\rangle$. More generally, a factorization of L into k irreducible factors translates into a chain of submodules

$$\{0\} =: M_0 \subsetneq \cdots \subsetneq M_k := K[\partial]/\langle L\rangle$$

such that each of the quotient modules M_i/M_{i-1} $(i = 1, \ldots, k)$ is simple. The Jordan-Hölder theorem implies that if we have another chain

$$\{0\} =: N_0 \subsetneq \cdots \subsetneq N_\ell := K[\partial]/\langle L\rangle$$

with N_i/N_{i-1} simple $(i = 0, \ldots, \ell)$, then $k = \ell$ and there is a permutation $\pi : \{1, \ldots, k\} \to \{1, \ldots, k\}$ such that $M_i/M_{i-1} \cong N_{\pi(i)}/N_{\pi(i)-1}$ for $i = 1, \ldots, k$. In this sense, the factorization of an operator is unique.

Example 4.49

1. The operator $L = D^2 \in C(x)[D]$ admits, among others, two factorizations $L = DD$ and $L = (D + \frac{1}{x-1})(D - \frac{1}{x-1})$. Let M_1, N_1 be the submodules of $C(x)[D]/\langle L\rangle$ generated by $[D]$ and $[D - \frac{1}{x-1}]$, respectively. We then have

   $$0 \subsetneq M_1 \subsetneq C(x)[D]/\langle L\rangle \quad \text{and} \quad 0 \subsetneq N_1 \subsetneq C(x)[D]/\langle L\rangle.$$

   Moreover, according to the discussion above, we should have $M_1 \cong N_1$ and

   $$(C(x)[D]/\langle L\rangle)/M_1 \cong (C(x)[D]/\langle L\rangle)/N_1.$$

   Indeed, it can be checked (Exercise 3) that automorphisms are given by

   $$M_1 \to N_1 \quad [qD] \mapsto [(x-1)qD - q],$$
   $$(C(x)[D]/\langle L\rangle)/M_1 \to (C(x)[D]/\langle L\rangle)/N_1 \quad [q] \mapsto [\frac{1}{x-1}q],$$

   for $q \in C(x)$.
2. The operator $L = S^2 - x \in C(x)[S]$ is irreducible, because if it were reducible, it would have a right factor of order 1, and the algorithms of Sect. 2.6 can be used to check that this is not the case. It follows that $C(x)[S]/\langle S^2 - x\rangle$ has no nontrivial submodules, i.e., none of the one-dimensional $C(x)$-subspaces of $C(x)[S]/\langle S^2 - x\rangle$ are closed under application of S. □

In the commutative case, we have the relation $pq = \operatorname{lcm}(p, q)\gcd(p, q)$ for any two monic polynomials $p, q \in C[x]$. Since two distinct monic irreducible

polynomials $p, q \in C[x]$ cannot have a common factor, we have $pq = \operatorname{lcm}(p, q)$ for such polynomials. The situation in the noncommutative case is different. Here, we must distinguish the question of whether an operator $L = K[\partial]$ can be written as the product of two smaller operators from the question of whether we can write it as the least common left multiple of two smaller operators. Of course, the latter implies the former, since $L = \operatorname{lclm}(A_1, A_2)$ implies the existence of B_1, B_2 such that $L = B_1 A_1 = B_2 A_2$. The converse, however, is not true.

Example 4.50 The operator $L = (xD + 1)D \in C(x)[D]$ is evidently the product of two first order operators. However, it cannot be written as a least common left multiple of two first order operators. To see why, note that $\log(x)$ is a solution of L. If we had $L = \operatorname{lclm}(D - a, D - b)$ for some $a, b \in C(x)$, then L would have two C-linearly independent hyperexponential solutions. In a differential field containing these solutions as well as $\log(x)$, there would be three linearly independent solutions, which by Theorem 3.20 is impossible if $\operatorname{ord}(L) = 2$. □

For a given operator, we can ask whether it can be written as a least common left multiple of smaller operators. As we have seen in the example above, this may not be the case even if the operator is not irreducible. If it is the case that the operator can be broken into a least common left multiple of one or more irreducible operators, we call it completely reducible.

Definition 4.51 An operator $L \in K[\partial] \setminus \{0\}$ is called *completely reducible* if there are irreducible operators $P_1, \ldots, P_k \in K[\partial]$ such that $L = \operatorname{lc}(L) \operatorname{lclm}(P_1, \ldots, P_k)$. □

Note that the case $k = 1$ is not excluded, so that the terminology as introduced in Definition 4.51 has the somewhat odd-looking side effect that every operator which is irreducible in the sense of Definition 4.48 is completely reducible in the sense of Definition 4.51. In the language of modules, we have seen above that L is irreducible if and only if $K[\partial]/\langle L\rangle$ is a simple module, i.e., one that has no submodules other than $\{0\}$ and $K[\partial]/\langle L\rangle$. Complete reducibility of an operator also translates into a classical notion of module theory: L is completely reducible if and only if $K[\partial]/\langle L\rangle$ is semisimple. A module M is called semisimple if it is a direct sum of simple submodules, or, equivalently, if for every submodule U of M there is another submodule W of M such that $M = U \oplus W$. The connection is established in the following proposition.

Proposition 4.52 *$L \in K[\partial] \setminus K$ is completely reducible if and only if $K[\partial]/\langle L\rangle$ is semisimple.* □

Proof "⇒": Let $L = \operatorname{lclm}(P_1, \ldots, P_k)$ for some monic and pairwise distinct irreducible operators $P_1, \ldots, P_k \in K[\partial]$. We can assume that the set of these operators is chosen minimally in the sense that no P_i is a right divisor of $\operatorname{lclm}(P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_k)$. Since the P_i are irreducible, we then have $\operatorname{ord}(L) = \operatorname{ord}(P_1) + \cdots + \operatorname{ord}(P_k)$.

Consider the natural module homomorphism

$$h\colon K[\partial]/\langle L\rangle \to K[\partial]/\langle P_1\rangle \times \cdots \times K[\partial]/\langle P_k\rangle.$$

Because $\langle L\rangle = \langle P_1\rangle \cap \cdots \cap \langle P_k\rangle$, the map h is injective. Since h is also a K-linear map between two K-vector spaces of the same dimension, it is even an isomorphism. We therefore have

$$K[\partial]/\langle L\rangle = h^{-1}(K[\partial]/\langle P_1\rangle \times \{0\}^{k-1}) \oplus \cdots \oplus h^{-1}(\{0\}^{k-1} \times K[\partial]/\langle P_k\rangle).$$

Since the P_i are irreducible, the $K[\partial]/\langle P_i\rangle$ are simple, and since h is an isomorphism, their preimages are simple as well. We have therefore written $K[\partial]/\langle L\rangle$ as a direct sum of simple submodules, so $K[\partial]/\langle L\rangle$ is semisimple.

"$\Leftarrow$": For every submodule M of $K[\partial]/\langle L\rangle$ there is a $Q \in K[\partial] \setminus \{0\}$ of minimal order such that M is generated by $[Q]$ (Exercise 10). It is then generated as a K-vector space by $[Q], \ldots, [\partial^{\operatorname{ord}(L)-\operatorname{ord}(Q)-1} Q]$, and the linear dependence of $[Q], \ldots, [\partial^{\operatorname{ord}(L)-\operatorname{ord}(Q)} Q] \in K[\partial]/\langle L\rangle$ over K shows that there is a $P \in K[\partial]$ with $\operatorname{ord}(P) = \operatorname{ord}(L) - \operatorname{ord}(Q)$ such that $PQ = L$. This implies that $M \cong K[\partial]/\langle P\rangle$, and if M is simple, P is irreducible.

Suppose that $M_1, \ldots, M_k \subseteq K[\partial]/\langle L\rangle$ are simple submodules such that

$$K[\partial]/\langle L\rangle = M_1 \oplus \cdots \oplus M_k,$$

and let $P_1, \ldots, P_k$ be irreducible such that $M_i \cong K[\partial]/\langle P_i\rangle$ for $i = 1, \ldots, m$. Writing $[1] = m_1 + \cdots + m_k$ with $m_1 \in M_1, \ldots, m_k \in M_k$ shows that $\operatorname{lclm}(P_1, \ldots, P_k)$ annihilates $[1]$ and is thus a left multiple of L. Since $\operatorname{ord}(P_i) = \dim_K(M_i)$ for $i = 1, \ldots, k$ and $\operatorname{ord}(L) = \dim_K K[\partial]/\langle L\rangle = \dim_K(M_1) + \cdots + \dim_K(M_k)$, the order of $\operatorname{lclm}(P_1, \ldots, P_k)$ cannot exceed $\operatorname{ord}(L)$. We must therefore have $L = \operatorname{lc}(L) \operatorname{lclm}(P_1, \ldots, P_k)$, as claimed. ∎

So far, we have only discussed general properties of factorization in $K[\partial]$ but no algorithms for finding factors. In Sect. 1.4, we have remarked that there is no general factorization algorithm for the commutative ring $C[x]$ of univariate polynomials over a field C. Instead, the ground field C determines if and how we can factor polynomials. As the commutative case is a special case of the factorization problem in an Ore algebra $K[\partial]$, it is clear that we can also not expect a uniform factorization algorithm applicable to all Ore algebras. We must make certain assumptions on $K[\partial]$, some of which will be "with loss of generality".

Without loss of generality, it suffices to focus on right factors. The reason is that if $K[\partial]$ is an Ore algebra with certain maps $\sigma, \delta\colon K \to K$, we can associate to it the Ore algebra $K[\partial^*]$ with $\sigma^*\colon K \to K$ defined by $\sigma^* = \sigma^{-1}$ and $\delta^*\colon K \to K$ defined by $\delta^* = -\delta \circ \sigma^{-1}$. For every $P = p_0 + p_1\partial + \cdots + p_r\partial^r \in K[\partial]$ we can then define $P^* = p_0 + \partial^* p_1 + \cdots + (\partial^*)^r p_r \in K[\partial^*]$. The operator $P^* \in K[\partial^*]$ is called the *adjoint* of $P \in K[\partial]$. A key feature of the adjoint is that $(PQ)^* = Q^*P^*$, so it translates right factors to left factors and vice versa.

It is also without loss of generality that we can restrict our attention to Ore algebras with $\sigma = \mathrm{id}$ or $\delta = 0$. The reason is that if $K[\partial]$ is an Ore algebra with

certain maps $\sigma, \delta\colon K \to K$ with $\sigma \neq \mathrm{id}$, it is isomorphic to the Ore algebra $K[\tilde{\partial}]$ with σ and 0. As shown in Exercise 14, any choice $\alpha \in K$ with $\sigma(\alpha) \neq \alpha$ gives rise to an isomorphism $h\colon K[\tilde{\partial}] \to K[\partial]$ defined by

$$h(\tilde{\partial}) = \alpha\partial - \partial\alpha.$$

This operation is known as the *Hilbert twist.*

For an Ore algebra $K[\partial]$ with $\sigma = \mathrm{id}$ or $\delta = 0$, let $\theta\colon K \to K$ be equal to σ if $\sigma \neq \mathrm{id}$, and equal to δ otherwise. We will make the following assumptions throughout the rest of this section:

Assumption 4.53

1. $\sigma = \mathrm{id}$ or $\delta = 0$.
2. C is an algebraically closed field and for every $p \in C[x]$ with $\deg p > 0$ we can compute a root.
3. There is an algorithm which for given $p_0, \ldots, p_r \in K$, $p_r \neq 0$, computes a basis of the C-vector space of all $y \in K$ with $p_0 y + p_1\theta(y) + \cdots + p_r\theta^r(y) = 0$.
4. For any $L \in K[\partial]$ there is a module F such that L admits a solution space $V(L)$ in F with $\dim_C V(L) = \mathrm{ord}(L)$ and no operator has a solution space whose dimension exceeds its order.

The first assumption is justified by the Hilbert twist. The second assumption saves us from the trouble related to factorization in $C[x]$ which technically is included as a special case, but which is not really our business here. The third assumption is justified at least in the cases $C(x)[S]$ and $C(x)[D]$, by the techniques discussed in Sects. 2.5 and 3.5. The fourth assumption can also be justified for these algebras, using the Picard-Vessiot theory briefly sketched at the ends of Sects. 2.2 and 3.2. Note that the assumption is only that an appropriate module F exists, not that we can actually construct it or compute the solutions it is supposed to contain.

We first describe an algorithm which is relatively easy but is only guaranteed to succeed for completely reducible operators. This algorithm is known as the eigenring method. Let $L \in K[\partial]$ and $r = \mathrm{ord}(L)$. The idea of the eigenring method is to search for operators $P \in K[\partial]$ which commute with ∂ modulo L in the sense that we have $[P\partial] = [\partial P]$ in $K[\partial]/\langle L\rangle$. The commutation with ∂ ensures that any such operator P acts as a C-linear map on the solution space $V(L)$. Eigenvectors of this C-linear map are elements of $V(L)$ on which P acts like a multiplication by a constant λ, the corresponding eigenvalue. This means that $P - \lambda$ annihilates the eigenvectors, so these eigenvectors are common solutions of $P - \lambda$ and L and hence of $\mathrm{gcrd}(P - \lambda, L)$. If we arrange that $0 < \mathrm{ord}(P) < \mathrm{ord}(L)$, then this greatest common right divisor will be a nontrivial right factor of L.

An operator P which commutes with ∂ modulo L amounts to a C-linear map from $V(L)$ to itself. It is clear that every element of the class $[P] \in K[\partial]/\langle L\rangle$ amounts to the same map. In particular, P commutes with ∂ modulo L if and only if $\mathrm{rrem}(P, L)$ commutes with ∂ modulo L. Moreoever, the set of all operators P which commute with ∂ modulo L is closed under addition and multiplication (Exercise 12)

and therefore forms a subring of $K[\partial]$. Then the subset of $K[\partial]/\langle L\rangle$ consisting of all classes $[P]$ with $[P\partial] = [\partial P]$ also forms a ring together with the operations $[P] + [Q] := [P + Q]$ and $[P][Q] := [PQ]$. This ring is called the *eigenring* of L and denoted by E_L.

Since an operator P which commutes with ∂ modulo L maps solutions of L to solutions of L, we must have $\operatorname{rrem}(LP, L) = 0$ for any such P. Conversely, if P is such that $\operatorname{rrem}(LP, L) = 0$, then LP annihilates all elements of $V(L)$, so P maps solutions of L to solutions of L. It is thus a C-linear map from $V(L)$ to itself and therefore commutes with ∂ modulo L. We have now shown that $[P]$ is an element of the eigenring of L if and only if $\operatorname{rrem}(LP, L) = 0$, and we can use the latter condition to find elements of the eigenring by making an ansatz $P = p_0 + p_1\partial + \cdots + p_{r-1}\partial^{r-1}$ with undetermined coefficients $p_0, \ldots, p_{r-1} \in K$, computing $\operatorname{rrem}(LP, L)$ and equating the coefficients of powers of ∂ to zero. This leads to a coupled system of functional equations for the undetermined coefficients, whose solution space can be computed.

Once we have found an element $[P]$ of the eigenring, we have to find an eigenvalue of the corresponding linear map $V(L) \to V(L)$. This can be done in two ways. We can either construct the minimal polynomial of $[P]$ by finding a C-linear dependence between $[1], [P], [P^2], \ldots$ and compute a root of this polynomial. Alternatively, we can exploit the fact that the resultant of two operators is zero if and only if the operators have a nontrivial greatest common right divisor (cf. Sect. 4.2). For an indeterminate z, the resultant $\operatorname{res}(L, P - z)$ is an element of $K[z]$ whose roots in C are exactly the eigenvalues of P.

Algorithm 4.54 *(Eigenring method)*
Input: $L \in K[\partial]$ *for an Ore algebra meeting the requirements specified in Assumption 4.53.*
Output: *A proper right factor of L, or an error message.*

1 *Compute a C-vector space basis $\{[P_1], \ldots, [P_d]\}$ of the eigenring E_L.*
2 *If $d = 1$, return "failed".*
3 *Choose a basis element $[P_i]$ with $0 < \operatorname{ord}(P_i) < \operatorname{ord}(L)$.*
4 *Compute an eigenvalue λ of $[P_i]$.*
5 *Return $\operatorname{gcrd}(L, P_i - \lambda)$.*

Concerning lines 2 and 3, note that the eigenring always contains $[1]$, but that this element is not useful because its eigenvalue is 1, so we would only get $\operatorname{gcrd}(L, 0) = L$ in step 5. As soon as $d > 1$, there must be a basis element with $0 < \operatorname{ord}(P_i) < \operatorname{ord}(L)$.

Example 4.55

1. Consider $L = (x-1)D^2 - x^2D + (x^2 - x - 1) \in C(x)[D]$. For computing the eigenring, we make an ansatz $P = p_0 + p_1D$ and enforce $\operatorname{rrem}(LP, L) = 0$.

$$\operatorname{rrem}(LP, L) = \left(-\frac{(x-2)xp_1}{x-1} - x^2p_0' - 2(1 - x + x^2)p_1' + (x-1)p_0''\right)$$

$$+\Big(\frac{(x-2)xp_1}{x-1}+2(x-1)p_0'+x^2p_1'+(x-1)p_1''\Big)D,$$

and equating the coefficients of D^0 and D^1 to zero gives a system of two functional equations for the two unknowns p_0, p_1. This system has the two linearly independent solutions $(1,0)$ and $(-\frac{1}{x-1}, \frac{1}{x-1})$. They give rise to the basis $\{[1], [-\frac{1}{x-1}+\frac{1}{x-1}D]\}$ of the eigenring E_L. We choose $P=-\frac{1}{x-1}(1-D)$ and compute an eigenvalue. Since $\operatorname{rrem}(P^2, L)=P$, the minimal polynomial is $z^2-z=(z-1)z$, so the eigenvalues are 0 and 1. Either of them leads to a right factor of L:

$$\operatorname{gcrd}(L, P-0)=D-1, \quad \operatorname{gcrd}(L, P-1)=D-x.$$

In fact, we have $L=\operatorname{lclm}(D-1, D-x)$.

2. Consider $L=D^2-(x+1)D+(x-1)\in C(x)[D]$. In this case, the ansatz $P=p_0+p_1D$ leads to

$$\begin{aligned}\operatorname{rrem}(LP, L)=&(-p_1+(-1-x)p_0'-2(-1+x)p_1'+p_0'')\\&+(p_1+2p_0'+(1+x)p_1'+p_1'')D,\end{aligned}$$

and equating the coefficients of D^0 and D^1 to zero gives a coupled system whose solution space turns out to be generated by $(p_0, p_1)=(1,0)$. Therefore, the algorithm aborts in line 2 with a failure. Note however that L admits the factorization $L=(D-1)(D-x)$. □

As long as no claim is made about the situations in which Algorithm 4.54 fails in finding a factor, it is obvious that the algorithm works as specified. Even an algorithm that trivially reports a failure for every input would be correct for this specification. The interesting feature of Algorithm 4.54 is that it is guaranteed to find a right factor whenever it is applied to an operator L which can be written as the least common left multiple of smaller operators.

Theorem 4.56 *Let* $U, W \in K[\partial]$ *be such that* $\operatorname{ord}(U), \operatorname{ord}(W) \geq 1$ *and* $\operatorname{grcd}(U, W)=1$. *Then Algorithm 4.54 applied to* $L=\operatorname{lclm}(U, W)$ *succeeds in finding a right factor.* □

Proof The algorithm only fails if the eigenring of L is generated by the class $[1]\in K[\partial]/\langle L\rangle$. (Keep in mind that we are assuming that C is algebraically closed, so there is no danger that there might not be any eigenvalue in line 4.) We show that this is not the case by exhibiting another element of the eigenring. By Theorem 4.21, there are $S, T\in K[\partial]$ such that $1=SU+TW$ and $\operatorname{ord}(S)<\operatorname{ord}(W)$, $\operatorname{ord}(T)<\operatorname{ord}(U)$. For $P=SU$ we have $0<\operatorname{ord}(P)<\operatorname{ord}(U)+\operatorname{ord}(W)=\operatorname{ord}(L)$. We show that $[P]$ is an element of the eigenring. Indeed, $\operatorname{rrem}(UP, U)=\operatorname{rrem}(USU, U)=0$ and $\operatorname{rrem}(WP, W)=\operatorname{rrem}(WSU, W)=\operatorname{rrem}(W(1-TW), W)=0$, so $LP=$

$\operatorname{lclm}(U, W)P$ contains both U and W as right factors. It therefore contains $L = \operatorname{lclm}(U, W)$ as a right factor, and we have shown $\operatorname{rrem}(LP, L) = 0$, as required. ■

As a corollary, if we know that L is completely reducible, then we can use Algorithm 4.54 to decide whether L is irreducible. This will be the case if and only if it fails to find a right factor. Algorithm 4.54 may also succeed with input that cannot be written as a least common left multiple. An example is the operator $(xD+1)D \in C(x)[D]$ from Example 4.50, for which it does find the right factor D although this operator is not a least common left multiple. In general, given an operator $L \in K[\partial]$ and a right factor U, it is not so obvious whether there is an operator $W \in K[\partial]$ with $\operatorname{gcrd}(U, W) = 1$ and $\operatorname{lclm}(U, W) = L$. If L is completely reducible, then the fact that $K[\partial]/\langle L\rangle$ is semisimple implies that for every right factor U of L there exists a suitable W. In general, it can happen that some right factors of L are part of an lclm and others are not. By the following theorem, whenever $L = AU$, then there exist B, W with $L = BW$ and $\operatorname{gcrd}(U, W) = 1$ if and only if there exists an S with $\operatorname{rrem}(US, A) = 1$. Testing this condition is similar to computing the eigenring.

Theorem 4.57

1. *If* $U, W, S, T, A, B \in K[\partial]$ *are such that* $SU + TW = 1$ *and* $AU - BW = 0$, *then* $\operatorname{rrem}(US, A) = 1$.
2. *If* $U, S, A \in K[\partial]$ *are monic and such that* $\operatorname{rrem}(US, A) = 1$, *and if* $W \in K[\partial]$ *is such that* $S \cdot y$ *is a solution of* W *if and only if* y *is a solution of* A, *then* $AU = \operatorname{lclm}(U, W)$. □

Proof

1. Let $L = AU = BW$. We consider the various operators as linear maps between solution spaces. The map $SU + TW = 1$ acts as identity on $V(L)$ and the map TW is the zero map on $V(W) \subseteq V(L)$, so SU is a projection of $V(L)$ onto $V(W)$ and acts as the identity on the image $V(W)$. The operator U maps $V(L)$ to $V(A)$ and has $V(U) \subseteq V(L)$ as the kernel. It is surjective because $\dim \operatorname{im} R = \dim V(L)/\ker R = \operatorname{ord}(L) - \operatorname{ord}(U) = \operatorname{ord}(A) = \dim V(A)$. Because of $SU + TW = 1$ we have $V(L) = V(U) \oplus V(W)$, hence U is an isomorphism from $V(W) \subseteq V(L)$ to $V(A)$. Since SU acts as the identity on $V(W)$, it follows that S is an isomorphism from $V(A)$ to $V(W)$. But then US acts as the identity on $V(A)$, so $US - 1$ maps $V(A)$ to zero, so $US - 1$ is a left multiple of A. It follows that $\operatorname{rrem}(US, A) = 1$.
2. If $\operatorname{rrem}(US, A) = 1$, then US acts on $V(A)$ as the identity, and S is an isomorphism from $V(A)$ to $S \cdot V(A)$. By assumption, $S \cdot V(A) = V(W)$. As US acts on $V(A)$ as the identity, U is an isomorphism from $V(W)$ to $V(A)$ and SU acts as the identity on $V(W)$. We therefore have $SU + TW = 1$ for a certain $T \in K[\partial]$, and hence $\operatorname{gcrd}(U, W) = 1$. It also follows that for every $y \in V(W)$ we have $U \cdot y \in V(A)$, i.e., $AU \cdot y = 0$. This means that W is a right factor of AU. Obviously, U is also a right factor of AU, so altogether $\operatorname{lclm}(U, W)$ is a right factor of AU. Since $V(W)$ and $V(A)$ are isomorphic,

we have $\operatorname{ord}(W) = \operatorname{ord}(A)$, and taking also $\operatorname{gcrd}(U, W) = 1$ into account, we have $\operatorname{ord}\operatorname{lclm}(U, W) = \operatorname{ord}(U) + \operatorname{ord}(W) = \operatorname{ord}(U) + \operatorname{ord}(A) = \operatorname{ord}(AU)$. Therefore, $\operatorname{lclm}(U, W) = pAU$ for some $p \in K$, but since A and U are monic by assumption, $p = 1$. ∎

Unless we know that the input is completely reducible, the eigenring method cannot be used for deciding whether a given operator is irreducible or not. The algorithm explained next, which is known as Beke's algorithm, can solve the factorization problem completely. In addition to the assumptions on $K[\partial]$ imposed in Assumption 4.53, we now include the following further assumptions:

Assumption 4.58

1. There is an algorithm which for any given $L \in K[\partial]$ finds all of its first order right factors.
2. There is an algorithm for solving systems of polynomial equations in C.
3. If $\sigma = \mathrm{id}$, $\delta \neq 0$, then for every operator $L \in K[\partial]$ there is an extension field E of K with $\operatorname{Const}(E) = \operatorname{Const}(K)$ such that $V(L) \subseteq E$ has dimension $\operatorname{ord}(L)$.
4. If $\sigma \neq \mathrm{id}$, $\delta = 0$, then for every operator $L \in K[\partial]$ which is not a left multiple of ∂ there is an extension ring E of K like in Theorem 2.27, with $\operatorname{Const}(E) = \operatorname{Const}(K)$, and such that $V(L) \subseteq E$ has dimension $\operatorname{ord}(L)$.

The last two assumptions allow us to formulate Wronskians. Recall that we defined $\theta = \sigma$ if $\delta = 0$ and $\theta = \delta$ if $\sigma = \mathrm{id}$. The *Wronskian* of some elements $y_0, \ldots, y_{r-1} \in E$ is defined as the determinant

$$W(y_0, \ldots, y_{r-1}) := \begin{vmatrix} y_0 & y_1 & \cdots & y_{r-1} \\ \theta(y_0) & \theta(y_1) & \cdots & \theta(y_{r-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \theta^{r-1}(y_0) & \theta^{r-1}(y_1) & \cdots & \theta^{r-1}(y_{r-1}) \end{vmatrix}.$$

Since θ is C-linear, it is clear that $W(y_0, \ldots, y_{r-1})$ is zero whenever $y_0, \ldots, y_{r-1}$ are C-linearly dependent. Conversely, if $y_0, \ldots, y_{r-1}$ are linearly independent over C and belong to the solution space $V(L) \subseteq E$ of some operator $L \in K[\partial]$, then $W(y_0, \ldots, y_{r-1})$ is nonzero. In this case, the Wronskian additionally satisfies a first order equation with coefficients in K. For differential equations, all of these facts are commonly covered in introductory courses, and the general assumptions declared above are chosen in such a way that the facts extend to the more general situation we consider here (Exercise 21).

Consider an operator $L = \ell_0 + \cdots + \ell_{r-1}\partial^{r-1} + \partial^r \in K[\partial]$ and let $y_0, \ldots, y_{r-1}$ be a basis of its solution space $V(L) \subseteq E$. If y is any element of $V(L)$, the elements $y, y_0, \ldots, y_{r-1}$ are linearly dependent, so their Wronskian is zero. Expanding the determinant along the first column, we get the equation

$$W_0 y - W_1 \theta(y) \pm \cdots + (-1)^r W_r \theta^r(y) = 0,$$

with

$$W_i = \begin{vmatrix} y_0 & y_1 & \cdots & y_{r-1} \\ \vdots & \vdots & \ddots & \vdots \\ \theta^{i-1}(y_0) & \theta^{i-1}(y_1) & \cdots & \theta^{i-1}(y_{r-1}) \\ \theta^{i+1}(y_0) & \theta^{i+1}(y_1) & \cdots & \theta^{i+1}(y_{r-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \theta^{r}(y_0) & \theta^{r-1}(y_1) & \cdots & \theta^{r}(y_{r-1}) \end{vmatrix} \qquad (i = 0, \ldots, r).$$

Since W_r is the Wronskian, which is not zero, we can divide the equation by $(-1)^r W_r$ to obtain a monic operator in $E[\partial]$ of order r whose solution space is equal to $V(L)$. If this operator were different from L, its greatest common right divisor with L would be an operator of order less than r with an r-dimensional solution space. Since this is impossible, we must have $\ell_i = (-1)^{r-i}\frac{W_i}{W_r}$ for $i = 0, \ldots, r-1$. In particular, the W_i are certain K-multiples of the Wronskian, and as such, they also satisfy certain first order equations.

If $L \in K[\partial]$ has a nontrivial right factor, then the coefficients of the right factor can also be expressed in terms of Wronskians. In particular, they satisfy certain first order equations. The key idea of the factorization algorithm is to generate from the input operator L some auxiliary equations which have these Wronskians as solutions. Since we assume that there is a way to find first order right factors, we can then determine candidates for the coefficients. Like in Sects. 2.6 and 3.6, these candidates will in general involve some undetermined constant parameters. By dividing L by such a parameterized candidate and forcing the remainder to zero, we finally get a system of polynomial equations for the parameters, which we can solve by assumption. The solutions give rise to the desired right factors.

It remains to be explained how to find suitable auxiliary equations. Suppose the input operator is $L = \ell_0 + \cdots + \ell_{r-1}\partial^{r-1} + \partial^r \in K[\partial]$, and let $s \in \{2, \ldots, r-1\}$. We seek right factors of order s, i.e., a factorization

$$L = Q \underbrace{(p_0 + \cdots + p_{s-1}\partial^{s-1} + \partial^s)}_{:=P}$$

with $Q \in K[\partial]$ and $p_0, \ldots, p_{s-1} \in K$. Let $y_0, \ldots, y_{s-1}$ be a basis of $V(P)$ and $W_0, \ldots, W_s \in E$ be the corresponding determinants as introduced above (with P playing the role of L). As determinants depend polynomially on their entries and the entries are defined in terms of L, we could use closure properties to construct for each of the determinants an annihilating operator. This would be a brutal approach. A somewhat less brutal way (but still quite costly) is to also exploit in the search that we need the $W_0, \ldots, W_s \in E$ to be pairwise similar in the sense that $W_i / W_s \in K$ for all i. This can be done by considering the K-subspace of E generated by all $s \times s$ determinants

$$\begin{vmatrix} \theta^{i_1}(y_0) & \cdots & \theta^{i_1}(y_{s-1}) \\ \vdots & \ddots & \vdots \\ \theta^{i_s}(y_0) & \cdots & \theta^{i_s}(y_{s-1}) \end{vmatrix} \in E$$

with $0 \le i_1 < \cdots < i_s < r$. There are $n := \binom{r}{s}$ many of these; let us call them Wronskian-type determinants and denote them by $\Delta_1, \ldots, \Delta_n$. Note that they include the determinants $W_0, \ldots, W_{s-1}$ we are interested in. It turns out that the K-vector space generated by $\Delta_1, \ldots, \Delta_n$ is closed under θ (Exercise 23), so there is a matrix $A \in K^{n \times n}$ with

$$\begin{pmatrix} \theta(\Delta_1) \\ \vdots \\ \theta(\Delta_n) \end{pmatrix} = A \begin{pmatrix} \Delta_1 \\ \vdots \\ \Delta_n \end{pmatrix}.$$

This matrix can be determined using only the input operator L, and the system can be solved using the techniques developed in the previous section.

Example 4.59

1. Consider $L = (x-1) + (x^2-1)S - xS^2 - (x-3)S^3 + S^4 \in C(x)[S]$. We want to decide if L has a right factor P of order $s = 2$. For a basis y_0, y_1 of $V(P)$, we consider the $\binom{4}{2} = 6$ determinants

$$\Delta_1 = \begin{vmatrix} y_0 & y_1 \\ \theta(y_0) & \theta(y_1) \end{vmatrix}, \quad \Delta_2 = \begin{vmatrix} y_0 & y_1 \\ \theta^2(y_0) & \theta^2(y_1) \end{vmatrix}, \quad \Delta_3 = \begin{vmatrix} y_0 & y_1 \\ \theta^3(y_0) & \theta^3(y_1) \end{vmatrix},$$

$$\Delta_4 = \begin{vmatrix} \theta(y_0) & \theta(y_1) \\ \theta^2(y_0) & \theta^2(y_1) \end{vmatrix}, \quad \Delta_5 = \begin{vmatrix} \theta(y_0) & \theta(y_1) \\ \theta^3(y_0) & \theta^3(y_1) \end{vmatrix}, \quad \Delta_6 = \begin{vmatrix} \theta^2(y_0) & \theta^2(y_1) \\ \theta^3(y_0) & \theta^3(y_1) \end{vmatrix}.$$

   The potential factor P has the form $P = \frac{\Delta_4}{\Delta_1} - \frac{\Delta_2}{\Delta_1}S + S^2$. Since θ is an automorphism, it commutes with the determinant, so we get $\theta(\Delta_1) = \Delta_4$, $\theta(\Delta_2) = \Delta_5$, and $\theta(\Delta_4) = \Delta_6$ for free. For the remaining Δ_i, we first apply θ and then use $\theta^4(y_i) = (x+3)\theta^3(y_i) + x\theta^2(y_i) - (x^2-1)\theta(y_i) + (1-x)y_i$ to obtain $\theta(\Delta_3) = (x+3)\Delta_5 + x\Delta_4 - (1-x)\Delta_1$, $\theta(\Delta_5) = (x+3)\Delta_6 + (x^2-1)\Delta_4 - (1-x)\Delta_2$, $\theta(\Delta_6) = -x\Delta_6 + (x^2-1)\Delta_5 - (1-x)\Delta_3$. We can put these relations together into the system

$$\begin{pmatrix} \theta(\Delta_1) \\ \theta(\Delta_2) \\ \theta(\Delta_3) \\ \theta(\Delta_4) \\ \theta(\Delta_5) \\ \theta(\Delta_6) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1-x & 0 & 0 & x & x+3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1-x & 0 & x^2-1 & 0 & x+3 \\ 0 & 0 & 1-x & 0 & x^2-1 & -x \end{pmatrix} \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \Delta_4 \\ \Delta_5 \\ \Delta_6 \end{pmatrix}.$$

Up to constant multiples, this system has exactly one solution whose components $\Delta_1, \Delta_2, \Delta_4$ are pairwise similar hypergeometric terms. This solution is

$$(\Delta_1, \ldots, \Delta_6) = (-1)^x\,(1, x+1, x^2+3x+3, -1, -x-2, 1)$$

and gives rise to the candidate $P = \frac{\Delta_4}{\Delta_1} - \frac{\Delta_2}{\Delta_1}S + S^2 = -1 - (x+1)S + S^2$, which indeed is a right factor of L.

2. Now consider $L = S^4 - 2(x+1)S^3 + (x^2+x-2)S^2 + 2xS + 1 \in C(x)[S]$. With $\Delta_1, \ldots, \Delta_6$ defined as before, we now get the system

$$\begin{pmatrix} \theta(\Delta_1) \\ \theta(\Delta_2) \\ \theta(\Delta_3) \\ \theta(\Delta_4) \\ \theta(\Delta_5) \\ \theta(\Delta_6) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -x^2-x+2 & 2(x+1) & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2x & 0 & 2(x+1) \\ 0 & 0 & 1 & 0 & 2x & x^2+x-2 \end{pmatrix} \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \Delta_4 \\ \Delta_5 \\ \Delta_6 \end{pmatrix}.$$

In this case, the solutions with hypergeometric components form a three-dimensional vector space generated by

$$b_1 = (-1)^x\,(0, -1, -2x, 0, 1, 0),$$
$$b_2 = (-1)^x\,(1, x, x^2+x+1, -1, -x-1, 1),$$
$$b_3 = (-1)^x\,(x, x^2, x^3+x^2+x+1, -x-1, -x^2-2x-1, x+2).$$

These solutions give rise to the parameterized family

$$P = S^2 - \frac{c_3x^2 + xc_2 - c_1}{c_2 + c_3x}S - \frac{c_3x + c_3 + c_2}{c_2 + c_3x}$$

of candidates, where c_1, c_2, c_3 are undetermined elements of C. In order to find out which elements of the family really are right factors, we compute

$$\operatorname{rrem}(L, P) = \frac{c_1^2 - c_2c_1 - c_3c_1 - c_3^2}{(c_3x + c_2)(c_3x + c_2 + 2c_3)} - \frac{(c_1^2 - c_2c_1 - c_3c_1 - c_3^2)(-c_3x^2 - c_2x + c_1)}{(c_3x + c_2)(c_3x + c_2 + c_3)(c_3x + c_2 + 2c_3)}S.$$

Since P is a right factor of L if and only if this expression is zero, the right factors are those with $c_1^2 - c_2c_1 - c_3c_1 - c_3^2 = 0$ and $(c_2, c_3) \neq (0, 0)$. □

Beke's factorization algorithm can be summarized as follows.

Algorithm 4.60 *(Beke)*
Input: $L \in K[\partial] \setminus \{0\}$, $s \in \{2, \dots, \operatorname{ord}(L) - 1\}$, *for an Ore algebra meeting the requirements of Assumptions 4.53 and 4.58.*
Output: *All right factors of L of order s.*

1 *Let* $n = \binom{\operatorname{ord}(L)}{s}$ *and write* $\Delta_1, \dots, \Delta_n$ *for the* $s \times s$ *minors of the matrix* $((\theta^i(y_j)))_{i=0,j=0}^{r-1,s-1}$, *where* $y_0, \dots, y_{s-1}$ *are place holders for some* C*-linearly independent solutions of L. Let* $i_0, \dots, i_s \in \{1, \dots, r\}$ *be the indices such that* Δ_{i_k} *is the minor obtained from the rows* $0, \dots, k-1, k+1, \dots, s$.
2 *Construct a matrix* $A \in K^{n \times n}$ *such that* $(\theta(\Delta_i))_{i=1}^n = A(\Delta_i)_{i=1}^n$.
3 *Find all solutions of the system from line 2 for which* $\Delta_{i_s} \neq 0$ *and the quotients* $\Delta_{i_j}/\Delta_{i_s}$ $(j = 0, \dots, s)$ *are in* K.
4 *The result of line 3 is a finite union of finite dimensional* C*-vector spaces. For each of these spaces, do the following:*
5 *Let* $b_1, \dots, b_d$ *be a basis and consider a generic element* $c_1 b_1 + \cdots + c_d b_d$ *with undetermined coefficients* $c_1, \dots, c_d$.
6 *Make an ansatz* $P = \sum_{j=0}^{s} (-1)^j \frac{\Delta_{i_j}}{\Delta_{i_s}} \partial^j$ *with the* Δ_i*'s replaced by the respective component of the generic element from line 5.*
7 *Compute* $\operatorname{rrem}(L, P)$ *and compute all* $(c_1, \dots, c_d) \in C^d$ *for which this remainder becomes zero. This may require solving a system of nonlinear equations. For every solution (or for an appropriate description of the solution set, if it is infinite), report the right factor* P.

Observe that in the computations for finding right factors, we never need to know any of the solutions $y_0, y_1, \dots$ explicitly. All we really need is that they are solutions and that they are linearly independent over C. Instead of assuming certain particular solutions, we may as well do the computations with appropriate formal objects. Multilinear algebra offers such formal objects. Consider the module $M = K[\partial]/\langle L \rangle$ and recall that finding a right factor of L of order s is the same as finding a submodule N of M with $\dim_K(N) = s$. The *exterior power* $\bigwedge^s V$ of a K-vector space V is a construction similar to the tensor product. It consists of formal objects that are written $v_1 \wedge \cdots \wedge v_s$ with $v_1, \dots, v_s \in V$, and of finite sums of such objects. Like the tensor product, these objects satisfy the rule

$$\begin{aligned} &v_1 \wedge \cdots \wedge v_{i-1} \wedge (p v_i + q v_i') \wedge v_{i+1} \wedge \cdots \wedge v_s \\ &= p\,(v_1 \wedge \cdots \wedge v_{i-1} \wedge v_i \wedge v_{i+1} \wedge \cdots \wedge v_s) \\ &+ q\,(v_1 \wedge \cdots \wedge v_{i-1} \wedge v_i' \wedge v_{i+1} \wedge \cdots \wedge v_s) \end{aligned}$$

for any $v_1, \dots, v_s \in V$ and $p, q \in K$, and unlike tensor products, they satisfy the additional rule

$$v_1 \wedge \cdots \wedge v_s = \operatorname{sgn}(\pi)(v_{\pi(1)} \wedge \cdots \wedge v_{\pi(s)})$$

for any $v_1, \ldots, v_s \in V$ and any permutation $\pi \in S_s$. If $b_1, \ldots, b_r$ form a basis of V, then a basis of $\bigwedge^s V$ is given by the objects $b_{i_1} \wedge \cdots \wedge b_{i_s}$ for all choices $i_1, \ldots, i_s \in \{1, \ldots, r\}$ with $i_1 < \cdots < i_s$. In particular, $\dim_K \bigwedge^s V = \binom{\dim(V)}{s}$.

The axioms for the exterior power $\bigwedge^s V$ are chosen in such a way that for any $P_1, \ldots, P_s \in K[\partial]$, the object $[P_1] \wedge \cdots \wedge [P_s] \in \bigwedge^s K[\partial]/\langle L \rangle$ can be interpreted as a determinant

$$\begin{vmatrix} P_1 \cdot y_1 & \cdots & P_1 \cdot y_s \\ \vdots & \ddots & \vdots \\ P_s \cdot y_1 & \cdots & P_s \cdot y_s \end{vmatrix}$$

where $y_1, \ldots, y_s$ are C-linearly independent solutions of L in some extension E of K. In accordance with this interpretation, we turn the K-vector space $\bigwedge^s K[\partial]/\langle L \rangle$ into a $K[\partial]$-module. Depending on whether $\theta = \sigma$ or $\theta = \delta$, the action of ∂ is defined through

$$\delta(v_1 \wedge \cdots \wedge v_s) = \sum_{i=1}^{s} (v_1 \wedge \cdots \wedge v_{i-1} \wedge \delta(v_i) \wedge v_{i+1} \wedge \cdots \wedge v_s),$$

$$\text{or } \sigma(v_1 \wedge \cdots \wedge v_s) = \sigma(v_1) \wedge \cdots \wedge \sigma(v_s).$$

Then $\bigwedge^s K[\partial]/\langle L \rangle$ corresponds to the K-vector space generated by all Wronskian-type determinants considered in Beke's algorithm, and the module structure imposed on it reflects the fact that this space is closed under applying θ.

It can be checked that whenever N is a submodule of $K[\partial]/\langle L \rangle$ with $\dim_K N = s$, then $\bigwedge^s N$ is (isomorphic to) a submodule of $\bigwedge^s K[\partial]/\langle L \rangle$. Since $\dim_K(\bigwedge^s N) = \binom{s}{s} = 1$, we are interested in the one-dimensional submodules of $\bigwedge^s K[\partial]/\langle L \rangle$. The search for such submodules corresponds to the search for solution vectors with hyperexponential/hypergeometric components of the coupled system in Beke's algorithm. Not every one-dimensional submodule of $\bigwedge^s K[\partial]/\langle L \rangle$ must be of the form $\bigwedge^s N$ for some submodule N of $K[\partial]/\langle L \rangle$. The one-dimensional submodules of interest are those which are generated by an element of the form $v_1 \wedge \cdots \wedge v_s$ (rather than by a certain K-linear combination of such terms). Searching within the set of all one-dimensional submodules for those of the required form corresponds to line 7 in Beke's algorithm where we compute the right remainder of L by a generic element involving parameters and solve a system of nonlinear equations in order to force the remainder to zero.

In conclusion, we do not lose anything by considering exterior powers instead of Wronskian-type determinants. We can however gain something by importing some general knowledge about exterior powers. In particular, the so-called Plücker relations can be used to substantially reduce the computational complexity of the algorithm. We do not discuss this here but refer to the literature for this optimization and other improvements.

Even when all known optimizations are applied, factorization of operators is extremely expensive. We have seen in Sects. 2.6 and 3.6 that finding hypergeometric and hyperexponential solutions involves a combinatorial search which may take an exponential amount of time. Here we apply these algorithms to operators of order $\binom{r}{s}$, and this binomial is itself exponential when r grows and r/s is approximately constant. In practice, this can mean that checking whether an operator of order 10 is irreducible might well be infeasible.

In order to show that a given operator of order r is irreducible, we can apply Beke's algorithm to search for right factors of order s, for any $s < r$. It is irreducible if and only if this search yields no results. In order to write a given operator of order r as a product of irreducible operators, we can use Beke's algorithm to find some right factor and then apply it recursively to factor the factors. To keep the growth of $\binom{r}{s}$ under control, it is a good idea to start with small factors, and to also exploit that finding a right factor of order s is the same as finding a left factor of order $r - s$. An implementation will roughly look as follows.

Algorithm 4.61
Input: $L \in K[\partial] \setminus \{0\}$ *satisfying Assumptions* 4.53 *and* 4.58.
Output: *A list* $(P_1, \ldots, P_m)$ *of irreducible elements of* $K[\partial]$ *such that* $L = P_1 \cdots P_m$.

if $\operatorname{ord}(L) = 1$ *then*
  Return (L).
Use the eigenring method (Algorithm 4.54) to search for a right factor P. *If it succeeds, apply the algorithm recursively to* $\operatorname{rquo}(L, P)$ *as well as to* P, *and return the concatenation of the resulting lists. Otherwise, continue as follows.*
for $s = 1, \ldots, \lfloor \operatorname{ord}(L)/2 \rfloor$ *do*
  Search for a right factor P *of* L *of order* s.
  if there is one then
    Recursively compute a factorization $(P_1, \ldots, P_{m-1})$ *of the right quotient* $\operatorname{rquo}(L, P)$ *and return* $(P_1, \ldots, P_{m-1}, P)$.
  else if $s \neq \operatorname{ord}(L)/2$ *then*
    Using the adjoint, search for a left factor P *of* L *of order* s.
    if there is one then
      Recursively compute a factorization $(P_2, \ldots, P_m)$ *of the left quotient* $\operatorname{lquo}(L, P)$.
      Return $(P, P_2, \ldots, P_m)$.
Return (P).

One reason why factorization of operators is more difficult than factorization in commutative polynomial rings $C[x]$ is that the solution set of an operator is a C-vector space while the roots of a polynomial $p \in C[x]$ are only finitely many. Since every root of $p \in C[x]$ must be a root of one of its factors, it is possible to design factorization algorithms for $C[x]$ based on the idea of approximating a root (for instance numerically) and then constructing a polynomial q of degree less than

$\deg(p)$ that has, up to the approximation error, the same root. Then $\gcd(p, q)$ is a factor of p, and by making the approximation accuracy sufficiently high and the coefficients of q sufficiently small (in a suitable measure), it can be ensured that $\gcd(p, q) = 1$ implies that p is irreducible.

For an operator L, say in the differential case, we can also select a solution, say in $C[[x]]$, compute it to a high accuracy, then use guessing (Sect. 1.5) to find a candidate for a lower order annihilating operator P of the solution, and, if we find one, compute $\mathrm{gcrd}(P, L)$ to obtain a right factor of L. The problem is that it is not obvious how to select the solution in the first place. The solutions that are annihilated by a right factor of L belong to a subspace of $V(L)$, and if we take an arbitrary element of $V(L)$, it is very unlikely that this element belongs to such a subspace.

We can increase our chances a bit by considering generalized series solutions at a singularity rather than power series solutions at an ordinary point. Suppose that $L = \mathrm{lclm}(P, Q)$ for some P, Q and assume that ξ is a non-apparent singularity of P but not of Q. Then every generalized series solution of Q at ξ is in fact a power series, but P must have at least one generalized series solution at ξ which is not a power series. Since the solutions of P are also solutions of L, we can take one of the generalized series solutions of L at ξ which is not a power series and use guessing to find an annihilating operator of order less than $\mathrm{ord}(L)$ for it. If we take sufficiently many terms of the generalized series into account, this computation is guaranteed to find an operator which has a nontrivial greatest common right divisor with L. Similarly, if $L = QP$ and there is a singularity ξ at which there are generalized series solutions with more than $\mathrm{ord}(Q)$ many different types, then there must be one type for which all generalized series solutions of L are already annihilated by P, because Q cannot have more than $\mathrm{ord}(Q)$ many linearly independent solutions.

Example 4.62 The operator

$$L = 36x^2D^4 + 144xD^3 + (36x^3 - 36x^2 + 9x + 80)D^2 + 18D + (x - 1)(9x + 8) \in C(x)[D]$$

has no first order right factors, and we want to know if there are any second order factors. The indicial polynomial of L at $\xi = 0$ is $\eta = 4x(x - 1)(3x - 2)(3x - 1)$, its roots $0, 1/3, 2/3, 1$ belong to three different $\mathbb{Z}$-equivalence classes. Therefore, if there is a factorization $L = QP$ with $\mathrm{ord}(Q) = \mathrm{ord}(P) = 2$, then for at least one of these classes, all of its corresponding generalized series solutions must be annihilated by P. For the exponent $1/3$, the only generalized series solution (up to constant multiples) is

$$x^{1/3}(1 - \tfrac{3}{8}x + \tfrac{9}{320}x^2 - \tfrac{9}{10240}x^3 + \tfrac{27}{1802240}x^4 + \cdots).$$

With a few more terms, guessing can find the candidate annihilating operator $P = 36x^2D^2 + (9x + 8)$, and we can easily check that P is indeed a right factor of L by computing $\mathrm{rrem}(L, P) = 0$. On the other hand, for the exponent 1, we have the

solution

$$x^1(1 - \tfrac{9}{80}x + \tfrac{563}{53760}x^2 - \tfrac{40907}{23654400}x^3 + \tfrac{1945123}{17220403200}x^4 + \cdots),$$

and even with hundreds of additional terms, we do not find any plausible candidates for annihilating operators of order 2. □

Exercises

1. Fix a $c \in C$ and let $L = (S - c)^2 \in C(x)[S]$. Determine all first order right factors of L.

2. The minimal polynomial of an algebraic function or algebraic number is always irreducible. Is it also true that the minimal order annihilating operator of a D-finite function is always irreducible?

3⋆. Check that the maps in Example 4.49 are indeed module isomorphisms.

4⋆. Let $L \in K[\partial]$ and $r = \operatorname{ord}(L)$. Show that L is irreducible if and only if every nonzero vector $p \in K^r$ is cyclic for the companion matrix C_L.

Hint: Use the isomorphism of Exercise 16 of Sect. 4.3.

5. Find an operator $L \in C(x)[S]$ which is reducible but not completely reducible.

6. Let K be a differential field and $L \in K[D]$. Let E be an extension of K such that $V(L) \subseteq E$ has dimension $\operatorname{ord}(L)$. Show that L is completely reducible when viewed as an element of $E[D]$.

7. Let M be a $K[\partial]$-module and consider a chain of submodules $\{0\} = M_0 \subsetneq \cdots \subsetneq M_k = M$ such that M_i/M_{i-1} is simple for every $i = 1, \ldots, k$. Prove or disprove: For every permutation $\pi \in S_k$ there exists a chain of submodules $\{0\} = N_0 \subsetneq \cdots \subsetneq N_k = M$ such that N_i/N_{i-1} is simple for every $i = 1, \ldots, k$ and $N_{\pi(i)} \cong M_i$ for every $i = 1, \ldots, k$.

8⋆. Prove or disprove:

a. If $L_1, L_2 \in K[\partial]$ are irreducible, then so is $L_1 \otimes L_2$.

b. If $P_1, P_2, P_3 \in K[\partial]$ are monic, irreducible, and pairwise distinct, then P_3 cannot be a right factor of $\operatorname{lclm}(P_1, P_2)$.

c. If $L \in C(x)[D]$ is completely reducible, then every formal power series solution of L is a C-linear combination of certain formal power series solutions of the right factors of L.

9. Show that $L \in C[\partial]$ is completely reducible if and only if it is squarefree.

10. Show that for every $M \subseteq K[\partial]/\langle L\rangle$ there is a $P \in K[\partial]$ such that M is generated by $[P]$.

11⋆. Write $C(x)[S]/\langle \operatorname{lclm}(S - x, S - x^2)\rangle$ as a direct sum of two submodules.

12⋆⋆. Show that the eigenring of an operator is indeed a ring.

13. (Manfred Buchacher) Show that if L is completely reducible and U is a right divisor of L, then U is completely reducible.

14⋆. Check that the Hilbert twist is an algebra isomorphism and that it turns δ into zero.

15⋆. Show that $(PQ)^* = Q^*P^*$ and $P^{**} = P$ for all $P, Q \in K[\partial]$.

16⋆⋆. Show that an irreducible operator $P \in C(x)[D]$ has either only algebraic solutions or only transcendental solutions (except 0).

Hint: First show that every algebraic function has an annihilating operator in $C(x)[D]$ which only has algebraic solutions.

17. Show that $\operatorname{gcld}(A, B) = \operatorname{gcrd}(A^*, B^*)^*$ for all $A, B \in K[\partial]$. Here, gcld refers to the greatest common left divisor, whose definition is analogous to Definition 4.19.

18. In the case $\sigma \neq \mathrm{id}$, $\delta = 0$, show that every factorization of a monic operator $L \in K[\partial]$ with $\operatorname{rrem}(L, \partial) = 0$ into irreducible factors contains one factor ∂. Show also that the corresponding statement is false in the case $\sigma = \mathrm{id}$, $\delta \neq 0$.

19. In the proof of Theorem 4.56 we used the fact that whenever $L = AU = BV$ for some operators $L, A, U, B, V \in K[\partial]$, then $\operatorname{lclm}(U, V)$ is a right divisor of L. Why is this true?

20. Write the following operators as least common left multiples:

a. $(1 - x^3)D^5 + 3x^2D^4 + (x^4 - 7x)D^3 + 3D^2 + (2x^5 - 2x^2)D + 2x^4 - 8x$;

b. $(x^3 - 3x - 2)D^4 + (-x^4 + x^3 - x + 1)D^3 + (-2x^4 + 2x^3 + 3x^2 + 4x - 1)D^2 + (x^5 - 2x^2 + 3x)D - x^4 + 2x - 3$;

c. $(x^3 + x^2 - 4x - 4)S^4 + (-x^4 - 2x^3 + 3x^2 + 3x - 1)S^3 + (-2x^4 - 5x^3 + 8x^2 + 20x + 5)S^2 + (x^5 + 4x^4 + 2x^3 - 5x^2 - 9x - 9)S - x^4 - 4x^3 - x^2 + 6x$;

d. $S^4 - S^2 + (2x + 3)S - x^2 - 2x$.

21⋆. Let $L = \partial^r - \ell_{r-1}\partial^{r-1} - \cdots - \ell_0 \in K[\partial]$ and let $y_0, \ldots, y_{r-1} \in E$ be a set of solutions.

a. Assuming $\sigma \neq \mathrm{id}$, $\delta = 0$, show that

$$\theta(W(y_0, \ldots, y_{r-1})) = \ell_0 W(y_0, \ldots, y_{r-1}).$$

b. Assuming $\sigma = \mathrm{id}$, $\delta \neq 0$, show that

$$\theta(W(y_0, \ldots, y_{r-1})) = \ell_{r-1} W(y_0, \ldots, y_{r-1}).$$

c. Show that $W(y_0, \ldots, y_{r-1})$ is nonzero whenever $y_0, \ldots, y_{r-1}$ are C-linearly independent.

22. Let $L \in C(x)[D]$, let g be an algebraic function, and let $M \in C(x)[D]$ be the minimal order operator such that for every solution f of L, the composition $f \circ g$ is a solution of M (cf. Theorem 3.29). Does irreducibility of L imply irreducibility of M?

23⋆. Show that the K-vector space generated by Δ_i is closed under θ.

24. Write the following operators as product of irreducible operators:
- **a.** $D^4 - xD^3 + (2-x)D^2 + (x^2 - x - 5)D - x^2 + 4x + 1$;
- **b.** $D^5 + D^4 - xD^3 + (x-4)D^2 + (2x-1)D - x^2 + x + 1$;
- **c.** $S^4 - 3S^2 + (2x+1)S - x^2$;
- **d.** $S^4 - xS^3 - xS^2 + (x^2 - 2)S - x^2 + 3x$.

25. Determine all second order right factors of $D^5 - D^3 \in C(x)[D]$.

26. In Algorithm 4.61, we assume that the subroutine for finding right factors just delivers one factor, but we have seen that we may find a parameterized family of factors in one stroke. How can Algorithm 4.61 be modified so as to take advantage of such a situation?

27. In view of $\binom{r}{s} = \binom{r}{r-s}$, what is the point of using adjoints in Algorithm 4.61 for finding left factors?

28. (Maximilian Jaroschek) Let $L \in K[\partial]$ and suppose there is a $P \in K[\partial]$ and a $k \in \mathbb{N}$ such that $L = P^k$.
- **a.** In the case $\sigma = \mathrm{id}$, show that there exists a $p \in K$ with $p = \mathrm{lc}(P)^k$.
- **b.** In the case $\delta = 0$, show that there exists a $p \in K$ with $p = ([\partial^0]P)^k$.

29⋆. In this section, we have discussed the factorization problem for Ore algebras $K[\partial]$ over a field K. The factorization problem also makes sense in Ore algebras $R[\partial]$ where R is just a ring, for example in $C[x][D]$. Show that $L = (4x-4)D^2 + (6x-4)D - 9$ is irreducible as element of $C[x][D]$ but not as element of $C(x)[D]$.

30⋆. In this section, we have discussed the factorization problem with respect to the multiplication of an Ore algebra. Alternatively, we could also consider the factorization problem with respect to the symmetric product, i.e., we could ask whether a given operator $L \in K[\partial]$ can be written as the symmetric product of some operators of lower order. Show that the operator $L = D^4 - 6D^3 + 11D^2 - 6D \in C(x)[D]$ can be factored in this sense.

References

The Jordan-Hölder theorem is more widely known for groups, a proof for its module version can be found in the book of Anderson and Fuller [31]. The connection between the Jordan-Hölder theorem and factorization of operators is nicely explained in a tutorial paper of Gomez-Torrecillas [219]. The discussion at the beginning of this section was inspired by this paper. A more direct proof of the essential uniqueness of a factorization is given in Ore's paper [344] (Thm. 1 in Chapter II). He also discusses complete reducibility (Sect. 2 of Chapter II), a concept that was first introduced by Loewy [315]. Loewy further showed that every operator can be written as a product of completely reducible operators. Such a factorization

is called a Loewy decomposition. Loewy decompositions in the case of several variables were studied by Schwarz [400].

Giesbrecht [213, 214] introduced the eigenring method to factor elements of the Ore algebra $K[\partial]$ with $\delta = 0$ and K a finite field. In a sense, it is an adaption of Berlekamp's factorization algorithm for $\mathbb{Z}_p[x]$ [204] to the noncommutative setting. The approach was extended to arbitrary Ore algebras $C(t)[\partial]$ over a finite constant field C by Giesbrecht and Zhang [216]. Caruso and Le Borgne [123] propose a faster version of the algorithm. Singer [409] applies the eigenring method in characteristic zero for determining, without factoring, whether or not a differential operator is irreducible. At the end of this paper, Singer gives an account on the historical development of the ideas. Van Hoeij [442] proposes an efficient algorithm for finding elements of the eigenring.

The eigenring method is also sketched in Sect. 4.2.2 of the book by van der Put and Singer [441], and in some lecture notes of Li [310]. These two sources also discuss Beke's algorithm, and were a great help for preparing this section.

Beke's algorithm is due to Beke [52] and was first formulated for differential operators. It was improved by various people, including Schwarz [399], Grigoriev [226], Wolf [460], Bronstein [112], and Tsarev [429, 430]. For general Ore algebras, the algorithm was described by Bronstein and Petkovšek [115].

The idea to separate factors by analyzing local solutions at the singularities has been turned into a complete factorization algorithm by van Hoeij [443]. As this algorithm is more heavily based on the notion of solutions than the eigenring method or Beke's algorithm, it does not directly extend to other Ore algebras. Only very recently, some progress on an analogous algorithm for the shift case has been reported [475].

Another recent result which is useful for both theory and algorithm development is an explicit degree bound for the right factors a linear differential operator can have [98].

4.5 Several Variables

For an Ore algebra $K[\partial]$ acting on a module F, it is natural to think of ∂ as something like a derivation, and to view the elements of F as univariate objects with a variable on which the derivation acts. If there are several variables, we may want to use several derivations. For instance, we could associate one partial derivative to each variable. Definition 4.1 already offers this freedom, because it starts from an arbitrary ring R and declares what it means for $R[\partial]$ to be an Ore algebra. If R itself is already an Ore algebra, the construction yields an Ore algebra with two ∂'s, and if we want, we can iterate further to obtain an Ore algebra with as many ∂'s as we like. In this way we can construct, for example, an Ore algebra $C[x, y][\partial_1][\partial_2]$ which acts on the ring $C[[x, y]]$ of formal power series in two variables x, y, with ∂_1, ∂_2 acting as the partial derivations in x, y, respectively. We could also consider an Ore

algebra $C[n,k][\partial_1][\partial_2]$ acting on the space $C^{\mathbb{N}\times\mathbb{N}}$ of sequences $((a_{n,k}))_{n,k=0}^{\infty}$ in two variables, with ∂_1, ∂_2 acting as shift operators with respect to n, k, respectively. It is also possible to construct Ore algebras with different types of operators, for instance letting ∂_1 be a derivation and ∂_2 be a shift operator gives an Ore algebra that acts on sequences of power series.

In general, two generators ∂_1, ∂_2 of an Ore algebra $R[\partial_1][\partial_2]$ need not commute. For example, if $R = C[x], \sigma_1 = \mathrm{id}, \delta_1 = \frac{d}{dx}, \sigma_2(f(x)) = f(qx), \delta_2 = 0$, where q is a nonzero constant, we have $\partial_1\partial_2 = q\partial_2\partial_1$. This situation is not typical, and by using the notation $R[\partial_1, \ldots, \partial_n]$ instead of $R[\partial_1]\cdots[\partial_n]$, we shall indicate that the ∂_i do commute with each other. Although this commutativity is not a formal requirement of the theory, it represents the most relevant situation in applications, and we will mostly restrict our attention to this case. The arithmetic of such a *multivariate Ore algebra* $R[\partial_1, \ldots, \partial_n]$ can be described by n endomorphisms $\sigma_1, \ldots, \sigma_n : R \to R$ and n maps $\delta_1, \ldots, \delta_n : R \to R$ such that δ_i is a σ_i-derivation for every i. We have the commutation rules $\partial_i\partial_j = \partial_j\partial_i$ and $\partial_i u = \sigma_i(u)\partial_i + \delta_i(u)$ for all $i, j = 1, \ldots, n$.

For better readability, we will write $D_x, D_y, D_z, \ldots$ for the elements of Ore algebras that behave like partial derivations with respect to $x, y, z, \ldots$. We shall assume that these commute with each other (e.g., $D_xD_y = D_yD_x$) and that partial derivations commute with all of the variables they are not responsible for (e.g., $D_x y = yD_x$). Similarly, we will write $S_x, S_y, S_z, \ldots$ for the shift operators that map $x, y, z, \ldots$ to $x+1, y+1, z+1, \ldots$, respectively. Also in this case, we have commutation rules like $S_xS_y = S_yS_x$ and $S_x y = yS_x$. Using this notation, the Ore algebras suggested above would be written as $C[x,y][D_x, D_y]$, $C[n,k][S_n, S_k]$, $C[x,n][D_x, S_n]$, respectively. Typically, we will consider Ore algebras $R[\partial_1, \ldots, \partial_n]$ with $R = C[x_1, \ldots, x_n]$ or $R = C(x_1, \ldots, x_n)$ where each ∂_i commutes with every x_j $(i \neq j)$. An example not matching this common pattern is the algebra $C(x)[D_x, S_x]$, in which we can shift as well as differentiate the variable.

In the univariate case, we have defined a function as D-finite if it has a nonzero annihilating operator. This definition is no longer useful in the multivariate setting. We use instead the criterion appearing in part 1 of Theorem 4.12 as a definition.

Definition 4.63 Let $A = K[\partial_1, \ldots, \partial_n]$ be an Ore algebra over a field K and let F be an A-module.

1. For $f \in F$, we call $\mathrm{ann}(f) = \{ L \in A : L \cdot f = 0 \}$ the *annihilator* of f (in A).
2. $f \in F$ is called *D-finite* (with respect to the action of A on F) if the K-vector space $A \cdot f = \{L \cdot f : L \in A\} \subseteq F$ has finite dimension. □

Example 4.64

1. If $\mathbb{Q}(x,y)[D_x, D_y]$ acts on the function space F of all bivariate meromorphic functions, then the element $\exp(x+y) \in F$ is D-finite. Its annihilator contains the operators $D_x - 1$ and $D_y - 1$, so the vector space $\mathbb{Q}(x,y)[D_x, D_y]\cdot\exp(x+y)$ consists of all $\mathbb{Q}(x,y)$-multiples of $\exp(x+y)$ and thus has dimension 1.

The element $f = x + \exp(x+y) \in F$ is also D-finite. Its annihilator contains the operators $(1-x)D_x^2 + xD_x - 1$ and $D_y^2 - D_y$, which can be used to rewrite each derivative $D_x^i D_y^j \cdot f$ as a $\mathbb{Q}(x, y)$-linear combination of f, $D_x \cdot f$, $D_y \cdot f$, $D_x D_y \cdot f$. We therefore have $\dim_{\mathbb{Q}(x,y)} \mathbb{Q}(x, y)[D_x, D_y] \cdot f \leq 4$. The actual dimension is smaller because there are additional relations. Another element of the annihilator is $xD_x + (1-x)D_y - 1$, and this element can be used to rewrite $D_x \cdot f$ and $D_x D_y \cdot f$ as linear combinations of f and $D_y \cdot f$, so these two functions also generate the $\mathbb{Q}(x, y)$-vector space $\mathbb{Q}(x, y)[D_x, D_y] \cdot f$. In fact, they form a basis and the dimension of the space is 2.

2. Let F be the set of germs of bivariate sequences, i.e., the set of all equivalence classes of sequences $a\colon \mathbb{N} \times \mathbb{N} \to C$ modulo the equivalence relation that identifies all sequences that differ on the vanishing set of a nonzero bivariate polynomial (cf. Definition 1.13). If we let the Ore algebra $C(n, k)[S_n, S_k]$ act on F, then the binomial coefficient $\binom{n}{k}$, viewed as an element of F, is D-finite. Its annihilator contains the operators $(1+n)-(1-k+n)S_n$ and $(k-n)+(k+1)S_k$, reflecting the identities

$$\binom{n+1}{k} = \frac{n+1}{n-k+1}\binom{n}{k} \quad \text{and} \quad \binom{n}{k+1} = \frac{n-k}{k+1}\binom{n}{k}.$$

These identities show that for each $i, j \in \mathbb{N}$, we can rewrite $S_n^i S_k^j \cdot \binom{n}{k}$ as a $C(n, k)$-multiple of $\binom{n}{k}$. Hence $\dim_{C(n,k)} C(n, k)[S_n, S_k] \cdot \binom{n}{k} = 1$.
The element $f = 1 + \binom{n}{k} \in F$ is also D-finite. Its annihilator contains the operators

$$(n-k+2)S_n^2 - (2n-k+3)S_n + (n+1) \quad \text{and}$$
$$(k+2)(1+2k-n)S_k^2 - (2k+2-n)(n+1)S_k - (k-n)(3+2k-n).$$

With these operators, every term $S_n^i S_k^j \cdot f$ can be rewritten as a linear combination of f, $S_n \cdot f$, $S_k \cdot f$, $S_n S_k \cdot f$, so $\dim_{C(n,k)} C(n, k)[S_n, S_k] \cdot f \leq 4$. Like in the previous example, the dimension is actually smaller. Because of the additional annihilating operator

$$(n-k+1)(1+2k-n)S_n + k(k+1)S_k + (n^2 - 3kn + k^2 - 2k - 1),$$

the vector space $C(n, k)[S_n, S_k] \cdot f$ is already generated by f and $S_k \cdot f$. In fact they form a basis.
The Stirling numbers of the second kind $S_2(n, k)$ are annihilated by the operator $S_n S_k - (k+1)S_k - 1 \in C(n, k)[S_n, S_k]$. With this operator, every element $S_n^i S_k^j \cdot S_2(n, k)$ can be rewritten as a $C(n, k)$-linear combination of terms $S_n^p \cdot S_2(n, k)$ and $S_k^q \cdot S_2(n, k)$ for various $p, q \in \mathbb{N}$, but these terms can be shown to be linearly independent, and since there are infinitely many, the Stirling numbers are not D-finite even though their annihilator is not empty.

3. Let F be the set of all germs of univariate sequences in $C(x)$, and let $C(x,n)[D_x, S_n]$ act on F. The sequence $(P_n(x))_{n=0}^{\infty}$ of Legendre polynomials, viewed as an element of F, is D-finite. This sequence is defined recursively by $P_0(x) = 1$, $P_1(x) = x$, and

$$(n+2)P_{n+2}(x) - (2n+3)xP_{n+1}(x) + P_n(x) = 0 \qquad (n \in \mathbb{N}).$$

This recurrence alone implies that the sequence $(P_n(x))_{n=0}^{\infty}$ is D-finite with respect to the action of $C(x,n)[S_n]$, in other words, it is D-finite in the sense of Chap. 2. With respect to the action of $C(x,n)[D_x, S_n]$, the recurrence only allows us to rewrite any term $D_x^i S_n^j \cdot P_n(x)$ as a linear combination of terms $D_x^k \cdot P_n(x)$ and $D_x^\ell S_n \cdot P_n(x)$ for various $k, \ell \in \mathbb{N}$. There are infinitely many of these terms, but there is an additional annihilating operator $(1-x^2)D_x + (n+1)S_n - (n+1)x$ of $P_n(x)$, so $P_n(x)$ and $S_n \cdot P_n(x)$ already generate the whole $C(x,n)$-vector space $C(x,n)[D_x, S_n] \cdot P_n(x)$. □

In these examples, we can see that the dimension of $K[\partial_1, \ldots, \partial_n] \cdot f$ is finite by finding, for each i, an annihilating operator containing ∂_i but none of the other generators. The following proposition says that this works in general. Informally speaking, a multivariate object is D-finite if and only if it is D-finite as a univariate object for each of the variables.

Proposition 4.65 *Let $A = K[\partial_1, \ldots, \partial_n]$ be an Ore algebra over a field K and let F be an A-module. An element $f \in F$ is D-finite if and only if* $\operatorname{ann}(f) \cap K[\partial_i] \neq \{0\}$ *for all $i = 1, \ldots, n$.* □

Proof "⇒": If f is D-finite, then $\dim_K(A \cdot f) =: d < \infty$, so for every $i = 1, \ldots, n$, the elements $f, \partial_i \cdot f, \cdots, \partial_i^d \cdot f \in F$ are linearly dependent over K. The linear dependence corresponds to a nonzero element of $\operatorname{ann}(f) \cap K[\partial_i]$.

"⇐": If $\operatorname{ann}(f) \cap K[\partial_i] \neq \{0\}$ for all $i = 1, \ldots, n$, we can let $r_i \in \mathbb{N}$ ($i = 1, \ldots, n$) be such that $\operatorname{ann}(f) \cap K[\partial_i]$ contains an operator of order r_i. Using these operators, every term $\partial_1^{p_1} \cdots \partial_n^{p_n} \cdot f$ ($p_1, \ldots, p_n \in \mathbb{N}$) can be rewritten into a K-linear combination of terms $\partial_1^{q_1} \cdots \partial_n^{q_n} \cdot f$ with $0 \le q_i < r_i$ ($i = 1, \ldots, n$). Since there are only finitely many of such terms, it follows that $\dim_K(A \cdot f) < \infty$, so f is D-finite. ■

In the univariate case, a D-finite series or sequence is uniquely determined by an annihilating operator and a finite number of initial terms. One of the motivations behind Definition 4.63 is to have the same feature also in the case of several variables. Indeed, as long as no trouble is caused by singularities, the generalization is very natural. For notational simplicity, let us consider the case of two variables. In the differential case, consider a left ideal $I \subseteq C(x,y)[D_x, D_y]$ such that $\dim_{C(x,y)} C(x,y)[D_x, D_y]/I$ is finite, and let $B_1, \ldots, B_r \in C(x,y)[D_x, D_y]$ be such that their equivalence classes form a vector space basis of $C(x,y)[D_x, D_y]/I$. A bivariate formal power series $a(x,y) = \sum_{n,k=0}^{\infty} a_{n,k} x^n y^k \in C[x,y]$ is called a *solution* of I if $I \subseteq \operatorname{ann} a(x,y)$. For every $i, j \in \mathbb{N}$, there are $u_1, \ldots, u_r \in C(x,y)$

such that $D_x^i D_y^j$ is equivalent modulo I to $u_1 B_1 + \cdots + u_r B_r$. For any coefficient $a_{i,j}$ of a solution $a(x, y)$ of I, we must therefore have

$$a_{i,j} = [x^i y^j]a(x, y) = \frac{1}{i!j!}\big(D_x^i D_y^j \cdot a(x, y)\big)\big|_{x=y=0}$$
$$= \frac{1}{i!j!}\big((u_1 B_1 + \cdots + u_r B_r) \cdot a(x, y)\big)\big|_{x=y=0}.$$

As long as the denominators of the u_ℓ or the B_ℓ do not vanish for $x = y = 0$, all of these coefficients are uniquely determined once we know the finitely many values $(B_\ell \cdot a(x, y))|_{x=0,y=0}$ $(\ell = 1, \ldots, r)$. Conversely, every choice of constants $(B_\ell \cdot a(x, y))|_{x=0,y=0}$ $(\ell = 1, \ldots, r)$ gives rise to a solution of I.

If some of the B_ℓ have denominators that vanish for $x = y = 0$ or there are some $i, j \in \mathbb{N}$ for which the corresponding u_ℓ have denominators that vanish for $x = y = 0$, then we say that $(0, 0)$ is a singular point of I. In this case, it can be difficult to determine which initial values give rise to power series solutions. Note that whether $(0, 0)$ is a singularity or not depends not only on the ideal I, but also on the choice of the basis $B_1, \ldots, B_\ell$.

In the shift case, the situation is similar. Given a left ideal $I \subseteq C(n, k)[S_n, S_k]$ for which the dimension of $C(n, k)[S_n, S_k]/I$ is finite, and operators $B_1, \ldots, B_r \in C(n, k)[S_n, S_k]$ whose equivalence classes form a basis of $C(n, k)[S_n, S_k]/I$, a bivariate sequence $(a_{n,k})_{n,k=0}^\infty$ is a solution of I if $I \subseteq \operatorname{ann}(a_{n,k})_{n,k=0}^\infty$. For every $i, j \in \mathbb{N}$, there are $u_1, \ldots, u_r \in C(n, k)$ such that $S_n^i S_k^j$ is equivalent modulo I to $u_1 B_1 + \cdots + u_r B_r$, so for any solution $(a_{n,k})_{n,k=0}^\infty$ we must have

$$a_{i,j} = (a_{n,k})_{n,k=0}^\infty\big|_{n=i,k=j} = (a_{n+i,k+j})_{n,k=0}^\infty\big|_{n=k=0}$$
$$= \big(S_n^i S_k^j \cdot (a_{n,k})_{n,k=0}^\infty\big)\big|_{n=k=0} = \big((u_1 B_1 + \cdots + u_r B_r) \cdot (a_{n,k})_{n,k=0}^\infty\big)\big|_{n=k=0}.$$

Like before, trouble arises only if the u_ℓ or the B_ℓ cannot be evaluated at $n = k = 0$ because of vanishing denominators. Such trouble is not uncommon.

Example 4.66

1. Consider the ideal

$$I = \langle nS_n - (n + k), S_k - (n + k)\rangle \subseteq C(n, k)[S_n, S_k].$$

The vector space $C(n, k)[S_n, S_k]/I$ is generated by the equivalence class of $B = 1 = S_n^0 S_k^0$. If there is no singularity trouble, every choice $a_{0,0} \in C$ can be uniquely extended to a solution $(a_{n,k})_{n,k=0}^\infty$ of I. However, there is singularity trouble: for every $i, j \in \mathbb{N}$, we have

$$S_n^i S_k^j - \frac{(n+k)(n+k+1)\cdots(n+k+i+j-1)}{n(n+1)\cdots(n+i-1)} \in I,$$

and an attempt to evaluate

$$a_{i,j} = \Big(\frac{(n+k)(n+k+1)\cdots(n+k+i+j-1)}{n(n+1)\cdots(n+i-1)} a_{n,k}\Big)\Big|_{n=k=0}$$

for some $i, j \in \mathbb{N}$ with $i > 0$ leads to a division by zero. We can only conclude that $a_{0,j} = 0$ for all $j > 0$. More generally, if we know all terms $a_{i,0}$ ($i \in \mathbb{N}$) of a solution, the formula above lets us compute all other terms. But there are infinitely many initial values.

The equivalence class of the operator S_n is also a basis of the vector space $C(n,k)[S_n, S_k]/I$. For every $i, j \in \mathbb{N}$ with $i > 0$ we have $S_n^i S_k^j - r_{i,j} S_n \in I$, where

$$r_{i,j} = \begin{cases} \frac{(n+k+1)(n+k+2)\cdots(n+k+i+j-1)}{(n+1)(n+2)\cdots(n+i-1)} & \text{if } i > 0, \\ n(n+k+1)(n+k+2)\cdots(n+k+i+j-1) & \text{if } i = 0 \text{ and } j > 0, \\ \frac{n}{k+n} & \text{if } i = j = 0. \end{cases}$$

The attempt to evaluate $a_{i,j} = (r_{i,j} a_{n+1,k})|_{n=k=0}$ will succeed for every $i, j \in \mathbb{N}$ except for $(i, j) = (0, 0)$. Therefore, a sequence solution of I is uniquely determined by the single initial value $a_{1,0}$ and the isolated exceptional term $a_{0,0}$.

2. Consider the ideal $I = \langle (k-n)S_n + (n-k+1), (k-n)S_k + (n-k-1)\rangle \subseteq C(n,k)[S_n, S_k]$. Again the vector space $C(n,k)[S_n, S_k]/I$ has dimension 1 and every nonzero element of it can serve as a basis. Again, some bases are better than others.

   Taking $B = 1$, we find that $S_n^i S_k^j - \frac{k+j-n-i}{k-n} \in I$, but evaluating $a_{i,j} = (\frac{k+j-n-i}{k-n} a_{n,k})|_{n=k=0}$ fails for every choice $i, j \in \mathbb{N}$ with $i \neq j$. Taking any other monomial $S_n^u S_k^v$ with $u \neq v$ as B solves the problem, because $S_n^i S_k^j - \frac{k+j-n-i}{k+v-n-u} S_n^u S_k^v \in I$ leads to evaluating $a_{i,j} = (\frac{k+j-n-i}{k+v-n-u} a_{n+u,k+v})_{n=k=0}$, which succeeds when $u \neq v$. □

The examples above were chosen so that it was possible to resolve the problems caused by singularities. This is not always possible, and it is not always easy to decide whether it is possible or not, especially if the vector space dimension is greater than one. Such issues are part of the price we have to pay for working with an Ore algebra $K[\partial_1, \ldots, \partial_n]$ defined over a field K. If we are in a situation where we are not willing to pay this price, we are led to the notion of holonomy (also known as holonomicity), a concept closely related but not equivalent to D-finiteness. Its definition is motivated by Proposition 4.65.

Definition 4.67 Let $A = C[x_1, \ldots, x_n][\partial_1, \ldots, \partial_n]$ be an Ore algebra, let I be a left ideal of A, and let f be an element of an A-module F.

1. I is called *holonomic* if for every subset $U \subseteq \{x_1, \ldots, x_n, \partial_1, \ldots, \partial_n\}$ with $|U| = n + 1$ we have $I \cap C[U] \neq \{0\}$.
2. f is called *holonomic* if the ideal $\operatorname{ann}(f) = \{L \in A : L \cdot f = 0\} \subseteq A$ is holonomic. □

One feature of holonomy is that it can also describe "functions" that are zero almost everywhere. For example, an object $\delta(x, y)$ which is zero except when $x = y$ is annihilated by the operator $x - y \in C[x, y][D_x, D_y]$. A meromorphic function cannot be of this form, but suitably generalized notions of functions such as distributions may be. Note that we cannot formulate the annihilation of a nonzero object by $x - y$ in the Ore algebra $C(x, y)[D_x, D_y]$ because $x - y \in \operatorname{ann}(f)$ implies $\frac{1}{x-y}(x - y) = 1 \in \operatorname{ann}(f)$, which means $1 \cdot f = f = 0$.

Example 4.68

1. $x + \exp(xy^2) \in C[[x, y]]$ is holonomic with respect to the algebra $C[x, y][D_x, D_y]$ because its annihilator contains the operators

$$\begin{aligned}
L_1 &= xy^2 D_x^2 - D_x^2 - xy^4 D_x + y^4 && \in C[x, y][D_x] \subseteq C[x, y][D_x, D_y],\\
L_2 &= yD_y^2 - D_y - 2xy^2 D_y && \in C[x, y][D_y] \subseteq C[x, y][D_x, D_y],\\
L_3 &= D_y^3 - 4x^2 D_x D_y - 2x D_y && \in C[x][D_x, D_y] \subseteq C[x, y][D_x, D_y],\\
L_4 &= D_x^3 - y^2 D_x^2 && \in C[y][D_x, D_y] \subseteq C[x, y][D_x, D_y].
\end{aligned}$$

2. $\binom{n}{k}$ is holonomic with respect to $C[n, k][S_n, S_k]$ because its annihilator contains the operators

$$\begin{aligned}
L_1 &= (n - k + 1)S_n - (n + 1) && \in C[n, k][S_n],\\
L_2 &= (k + 1)S_k - (k - n) && \in C[n, k][S_k],\\
L_3 &= S_n S_k - S_k - 1 && \in C[n][S_n, S_k],\\
L_4 &= S_n S_k - S_k - 1 && \in C[k][S_n, S_k].
\end{aligned}$$

□

According to Proposition 4.65, an object is already D-finite if it has annihilating operators like the operators L_1 and L_2 in the examples above, and since holonomic objects must in addition have annihilating operators like L_3 and L_4, it seems that holonomy is a stronger requirement than D-finiteness. However, this is not necessarily the case. We will show next that if the Ore algebra is such that all σ_i are equal to id (like for example in the differential case), holonomy is in fact equivalent to D-finiteness. In the proof, we will construct the required operators by setting up a linear system with more variables than equations, as we have already done many times. This time however, we will use a linear system over C rather than over K.

Theorem 4.69 *Let $A = C[x_1, \dots, x_n][\partial_1, \dots, \partial_n]$ be an Ore algebra with $\sigma_1 = \cdots = \sigma_n = \mathrm{id}$. Let F be a $C[x_1, \dots, x_n][\partial_1, \dots, \partial_n]$-module which can also be viewed as a $C(x_1, \dots, x_n)[\partial_1, \dots, \partial_n]$-module, and let $f \in F$. Then f is holonomic if and only if it is D-finite.* □

Proof "⇒": If f is holonomic, then for every $i = 1, \dots, n$ there exists a nonzero annihilating operator in $C[x_1, \dots, x_n][\partial_i] \subseteq C(x_1, \dots, x_n)[\partial_i]$. From Proposition 4.65 it follows that f is D-finite.

"⇐": Write $K = C(x_1, \dots, x_n)$ and let $B = \{b_1, \dots, b_r\}$ be a basis of the K-vector space $V = K[\partial_1, \dots, \partial_n] \cdot f \subseteq F$. Without loss of generality, we may assume $b_1 = f$. To every element $g = u_1 b_1 + \cdots + u_r b_r$ of V we associate the coefficient vector $\hat{g} = (u_1, \dots, u_r) \in K^r$ with respect to B. There are matrices $A_1, \dots, A_n \in K^{r \times r}$ such that the coefficient vector of $\partial_i \cdot g$ is $A_i \hat{g} + \delta_i(\hat{g})$ ($i = 1, \dots, n$), where $\delta_i(\hat{g})$ means componentwise application of δ_i.

Let q be a common denominator of all entries of all A_i ($i = 1, \dots, n$), and let $d \geq 1$ be such that the total degree of q as well as the entries of the qA_i ($i = 1, \dots, n$) are less than d. For an element $g \in V$ with $\hat{g} = q^{-k}(p_1, \dots, p_r)$ for some $k \in \mathbb{N}$, and certain polynomials $p_1, \dots, p_r \in C[x_1, \dots, x_n]$ of total degree at most $u \in \mathbb{N}$, the coefficient vector of any $\partial_i \cdot g$ will have the form $q^{-(k+1)}(\tilde{p}_1, \dots, \tilde{p}_r)$ for certain polynomials $\tilde{p}_1, \dots, \tilde{p}_r \in C[x_1, \dots, x_n]$ of total degree at most $u + d$. (Here we used that $\sigma_i = \mathrm{id}$.) Also the coefficient vector of any $x_i \cdot g$ has this format. By induction, it follows that for every $k \in \mathbb{N}$ and every choice $i_1, \dots, i_n, j_1, \dots, j_n \in \mathbb{N}$ with $i_1 + \cdots + i_n + j_1 + \cdots + j_n \leq k$, the coefficient vector of

$$x_1^{i_1} \cdots x_n^{i_n} \partial_1^{j_1} \cdots \partial_n^{j_n} \cdot f \in V$$

has the form $q^{-k}(p_1, \dots, p_r)$ for certain polynomials $p_1, \dots, p_r \in C[x_1, \dots, x_n]$ of total degree at most kd.

Now let $U \subseteq \{x_1, \dots, x_n, \partial_1, \dots, \partial_n\}$ with $|U| = n + 1$. We have to show that f has an annihilating operator in $C[U]$. For a $k \in \mathbb{N}$, consider an ansatz $L = \sum_\tau c_\tau \tau$ for such an operator, where τ ranges over all terms $x_1^{i_1} \cdots x_n^{i_n} \partial_1^{j_1} \cdots \partial_n^{j_n}$ with $i_1 + \cdots + i_n + j_1 + \cdots + j_n \leq k$ but only involving variables from U (i.e., $i_\ell = 0$ if $x_\ell \notin U$ and $j_\ell = 0$ if $\partial_\ell \notin U$). Because of $|U| = n + 1$, the ansatz for L contains $\binom{n+1+k}{k}$ undetermined coefficients c_τ. By the analysis in the previous paragraph, the coefficient vector of $L \cdot f$ with respect to the basis B has the form $q^{-k}(p_1, \dots, p_r)$ for certain $p_1, \dots, p_r \in C[x_1, \dots, x_n]$ of total degree at most kd. Equating the coefficients of all the p_i ($i = 1, \dots, r$) with respect to the variables $x_1, \dots, x_n$ to zero yields a linear system over C with at most $r\binom{n+kd}{kd}$ equations.

For sufficiently large k, we have $\binom{n+1+k}{k} > r\binom{n+kd}{kd}$, because the left hand side is a polynomial in k of degree $n + 1$ (with a positive leading coefficient) while the right hand side is a polynomial in k of degree n. Therefore, when k is sufficiently large, the linear system will have a nonzero solution. This solution gives rise to the required nonzero operator L. ■

The restriction on the σ_i in the theorem above ensures that the rational functions appearing in the proof have a small common denominator. If the common denominator is small, the corresponding numerators are also small, and this means that coefficient comparison does not lead to too many equations. If some σ_i are different from id, as for example in the shift case, holonomy and D-finiteness are not the same.

Example 4.70 The rational function $a(n,k) = 1/(n^2+k^2)$ is D-finite because it satisfies the recurrence equations

$$(n^2+(k+1)^2)a(n,k+1) - (n^2+k^2)a(n,k) = 0,$$
$$((n+1)^2+k^2)a(n+1,k) - (n^2+k^2)a(n,k) = 0.$$

It is however not holonomic, because in order for $a(n,k)$ to be holonomic, we would also need an annihilating operator only containing n, S_k, S_n, i.e., a relation of the form

$$\sum_{u,v} \frac{p_{u,v}(n)}{(n+u)^2+(k+v)^2} = 0$$

for certain polynomials $p_{u,v}$, not all zero. Such a relation does not exist. To see why, suppose there is a pair (i,j) for which $p_{i,j}$ is not the zero polynomial. Then there is an $m \in \mathbb{Q} \setminus \mathbb{Z}$ such that $p_{i,j}(m) \neq 0$, so setting n to m in the relation above yields a C-linear dependence among the rational functions $1/((m+u)^2+(k+v)^2) \in C(k)$. But these rational functions are C-linearly independent because their denominators are pairwise coprime. See Exercise 8 for some more details. □

The lack of equivalence between holonomy and D-finiteness in the shift case can be a source of annoyance. For example, it can be shown (Exercise 5) that a sequence $(a_{n,k})_{n,k=0}^{\infty}$ is holonomic with respect to $C[n,k][S_n,S_k]$ if and only if its generating function $a(x,y) = \sum_{n,k=0}^{\infty} a_{n,k} x^n y^k$ is holonomic with respect to $C[x,y][D_x,D_y]$. The latter is equivalent to D-finiteness but, as we have just seen, the former is not. This means that Theorem 2.33 for translating recurrence equations to differential equations does not carry over to multivariate D-finite functions. Summation and integration also do not preserve D-finiteness in the case of several variables. For example, $\frac{1}{n+x}$ is D-finite with respect to $C(n,x)[S_n,D_x]$ but $\int \frac{1}{n+x}dx$ and $\sum_{k=1}^{n} \frac{1}{k+x}$ are not (Exercises 2 and 12).

Summation and integration are the subject of the next chapter. Other closure properties are less problematic. In particular, addition and (if meaningful) multiplication preserve D-finiteness.

Theorem 4.71 *Let $A = K[\partial_1,\ldots,\partial_n]$ be an Ore algebra and F be an A-module. Let $f,g \in F$ be D-finite.*

1. *$L \cdot f$ is D-finite for every $L \in A$.*
2. *$f+g$ is D-finite.*

3. *If* $m\colon F \times F \to F$ *is a* K*-bilinear function such that for every* $i = 1, \ldots, n$ *there are* $\alpha_i, \beta_i, \gamma_i \in K$ *with*

$$\partial_i \cdot m(u, v) = \alpha_i m(u, v) + \beta_i m(\partial_i \cdot u, v) + \beta_i m(u, \partial_i \cdot v) + \gamma_i m(\partial_i \cdot u, \partial_i \cdot v)$$

for all $u, v \in F$*. Then* $m(f, g)$ *is D-finite.* □

Proof

1. Since f is D-finite, $\dim_K(A \cdot f) < \infty$. $A \cdot (L \cdot f)$ is a subspace of $A \cdot f$ and therefore also has finite dimension. The claim follows.
2. Since f, g are D-finite, we have $\dim_K(A \cdot f), \dim_K(A \cdot g) < \infty$. Consequently, $\dim_K(A \cdot f + A \cdot g) < \infty$. The K-vector space $A \cdot (f + g)$ is a subspace of $A \cdot f + A \cdot g$ and therefore also has a finite dimension. The claim follows.
3. Since f, g are D-finite, we have $\dim_K(A \cdot f), \dim_K(A \cdot g) < \infty$. Consequently, $\dim_K((A \cdot f) \otimes_K (A \cdot g)) < \infty$. By the assumption on m, the K-vector space $A \cdot m(f, g)$ is isomorphic to (a subspace of) $(A \cdot f) \otimes_K (A \cdot g)$ and therefore also has a finite dimension. The claim follows. ■

Theorem 4.72

1. *Let* $f \in C[[x_1, \ldots, x_n]]$ *be D-finite (with respect to the algebra* $C(x_1, \ldots, x_n)[D_{x_1}, \ldots, D_{x_n}]$*) and suppose that elements* $g_1, \ldots, g_n \in C[[z_1, \ldots, z_m]]$ *are algebraic over* $C(z_1, \ldots, z_m)$ *but algebraically independent over* C*. If the composition* $f(g_1, \ldots, g_n)$ *is a well-defined element of* $C[[z_1, \ldots, z_m]]$*, then it is D-finite (w.r.t.* $C(z_1, \ldots, z_m)[D_{z_1}, \ldots, D_{z_m}]$*).*
2. *Let* $f\colon \mathbb{C}^n \to \mathbb{C}$ *be a meromorphic D-finite function (with respect to the algebra* $C(x_1, \ldots, x_n)[S_{x_1}, \ldots, S_{x_n}]$*). Suppose that linear functions* $g_1, \ldots, g_n\colon \mathbb{Q}^m \to \mathbb{Q}$ *are linearly independent over* $\mathbb{C}$*. Then the composition* $f(g_1, \ldots, g_n)$ *is D-finite (w.r.t.* $C(z_1, \ldots, z_m)[S_{z_1}, \ldots, S_{z_m}]$*).* □

Proof

1. The field $C(z_1, \ldots, z_m)(g_1, \ldots, g_n)$ is closed under application of D_{z_i} for every $i = 1, \ldots, m$ (for the same reason as in the univariate case), and it is a finite-dimensional $C(z_1, \ldots, z_m)$-vector space (also for the same reason as in the univariate case).
   Since f is D-finite, the functions $D_{x_1}^{e_1} \cdots D_{x_n}^{e_n} \cdot f$ form a finite-dimensional $C(x_1, \ldots, x_n)$-vector space. The substitution $x_i \mapsto g_i$ $(i = 1, \ldots, n)$ is well-defined on this space, because $g_1, \ldots, g_n$ are assumed to be algebraically independent over C, so plugging them into the denominator of an element of $C(x_1, \ldots, x_n)$ cannot produce a division by zero. Therefore, the functions $(D_{x_1}^{e_1} \cdots D_{x_n}^{e_n} \cdot f)(g_1, \ldots, g_n)$ form a finite-dimensional $C(g_1, \ldots, g_n)$-vector space, which we may also view as a $C(z_1, \ldots, z_m)(g_1, \ldots, g_n)$-vector space. Call this space V.
   For $h = f(g_1, \ldots, g_n)$, we have

$$D_{z_i} \cdot h = \sum_{j=1}^{n} \underbrace{\underbrace{\big((D_{x_j} \cdot f)(g_1, \dots, g_n)\big)}_{\in V} \overbrace{(D_{z_i} \cdot g_j)}^{\in C(z_1,\dots,z_m)(g_1,\dots,g_n)}}_{\in V} \in V$$

for every $i = 1, \dots, m$, by the chain rule. By induction, it follows that every $D_{z_1}^{e_1} \cdots D_{z_m}^{e_m} \cdot h$ belongs to V. Since V is a finite-dimensional vector space over $C(z_1, \dots, z_m)(g_1, \dots, g_n)$ and $C(z_1, \dots, z_m)(g_1, \dots, g_n)$ is a finite-dimensional vector space over $C(z_1, \dots, z_m)$, V is a finite-dimensional $C(z_1, \dots, z_m)$-vector space, and since V contains $C(z_1, \dots, z_m)[D_{z_1}, \dots, D_{z_m}] \cdot h$ as a subspace, we have shown that h is D-finite.

2. Using vector notation $z = (z_1, \dots, z_m)$, $g = (g_1, \dots, g_n)$, etc., we can write

$$S_{z_i} \cdot f(g(z)) = f(g(z) + g(e_i)) \qquad (i = 1, \dots, m),$$

where e_i is the ith unit vector in $\mathbb{C}^n$. More generally, for every choice $\ell_1, \dots, \ell_m \in \mathbb{N}$ we have

$$S_{z_1}^{\ell_1} \cdots S_{z_m}^{\ell_m} \cdot f(g(z)) = f(g(z) + \ell_1 g(e_1) + \cdots + \ell_m g(e_m)).$$

If $d \in \mathbb{N}$ is the common denominator of the entries of $g(e_1), \dots, g(e_m) \in \mathbb{Q}^n$, we find that each $S_{z_1}^{\ell_1} \cdots S_{z_m}^{\ell_m} \cdot f(g(z))$ belongs to the $C(x_1, \dots, x_n)$-vector space generated by $f(g(z) + u/d)$, where u runs through $\mathbb{N}^n$. Since f is D-finite, this space is also generated by $f(g(z) + u/d)$ where u runs through $\{0, \dots, r\}^n$ for some sufficiently large $r \in \mathbb{N}$. Hence, its dimension is finite.
The substitution $x_i \mapsto g_i$ $(i = 1, \dots, n)$ is well-defined on this space, because the assumption on $g_1, \dots, g_n$ implies that these functions are algebraically independent over C (Exercise 14), so plugging them into the denominator of an element of $C(x_1, \dots, x_n)$ cannot produce a division by zero. Therefore, the $C(z_1, \dots, z_m)$-vector space $C(z_1, \dots, z_m)[S_{z_1}, \dots, S_{z_m}] \cdot f(g)$ has finite dimension, which proves that $f(g)$ is D-finite. ∎

The condition of algebraic independence on the substitution arguments is required only for ensuring that there is no division by zero. If there is a dependence between the arguments, there is a good chance that the computation succeeds nevertheless, and it is worth giving it a try. If it fails, more advanced techniques discussed in the next chapter can be applied (cf. Theorems 5.30 and 5.36). In particular, these more advanced techniques are often needed when one of the arguments is set to a constant, e.g., when we want to compute a differential equation in x for $f(x, 0)$ from a system of differential equations in x, y for $f(x, y)$.

For the shift case, we have formulated the closure under substitution for meromorphic functions rather than sequences because this case can be formulated more conveniently. For sequences, we would have to restrict to substitutions that map (nonnegative) integer arguments to (nonnegative) integer arguments,

but the proof is otherwise the same. We have furthermore simplified matters by considering only the pure differential case and the pure shift case, respectively. The result can be extended to the mixed case, as long as it is ensured that the substitution only affects variables on which only derivations or shifts act. For example, given a D-finite element $f(x_1, x_2, n_1, n_2, u_1, u_2)$ of a module on which an Ore algebra $C(x_1, x_2, n_1, n_2, u_1, u_2)[D_{x_1}, D_{x_2}, S_{n_1}, S_{n_2}, \partial_{u_1}, \partial_{u_2}]$ acts, a substitution $x_1 = g_1(z_1, z_2)$, $x_2 = g_2(z_1, z_2)$, $n_1 = h_1(k_1, k_2)$, $n_2 = h_2(k_1, k_2)$ with g_1, g_2 algebraic and algebraically independent and h_1, h_2 linear and linearly independent yields a result which is D-finite with respect to the Ore algebra $C(z_1, z_2, k_1, k_2, u_1, u_2)[D_{z_1}, D_{z_2}, S_{k_1}, S_{k_2}, \partial_{u_1}, \partial_{u_2}]$.

Closure properties are easy to program if we do not represent a D-finite function f using annihilating operators but instead exploit that the K-vector space $K[\partial_1, \ldots, \partial_n] \cdot f$ has finite dimension. To have a finite dimension means that for some $r \in \mathbb{N}$ there is a K-vector space embedding $\phi\colon K[\partial_1, \ldots, \partial_n] \cdot f \to K^r$. Like in the proof of Theorem 4.69, we can turn K^r into a $K[\partial_1, \ldots, \partial_n]$-module and ϕ into a module homomorphism by choosing matrices $A_1, \ldots, A_n \in K^{r\times r}$ such that

$$\phi(\partial_i \cdot u) = A_i \sigma_i(\phi(u)) + \delta_i(\phi(u))$$

for $i = 1, \ldots, n$ and every $u \in K[\partial_1, \ldots, \partial_n] \cdot f$, where the applications of the σ_i and δ_i are meant componentwise. We call these matrices *companion matrices* for $\partial_1, \ldots, \partial_n$. Depending on the circumstances, it may be fair to say that we know the D-finite function f once we know the vector $\phi(f) \in K^r$, the companion matrices, and the functions $\sigma_1, \ldots, \sigma_n, \delta_1, \ldots, \delta_n$ defining the Ore algebra.

With this point of view, if f, g are two D-finite functions and we have module embeddings $\phi\colon K[\partial_1, \ldots, \partial_n] \cdot f \to K^r$ and $\psi\colon K[\partial_1, \ldots, \partial_n] \cdot g \to K^s$ with companion matrices $A_1, \ldots, A_n \in K^{r\times r}$ and $B_1, \ldots, B_n \in K^{s\times s}$, then we can encode $h := f + g$ by the embedding $\chi\colon K[\partial_1, \ldots, \partial_n] \cdot h \to K^{r+s}$ defined by $\chi(h) = \binom{\phi(f)}{\psi(g)} \in K^{r+s}$ and the companion matrices

$$\begin{pmatrix} A_1 & \\ & B_1 \end{pmatrix}, \ldots, \begin{pmatrix} A_n & \\ & B_n \end{pmatrix} \in K^{(r+s)\times(r+s)}.$$

This takes literally no computation time. For other closure properties, the construction is only slightly more involved.

If a D-finite function f is given in the sense outlined above, i.e., if for a certain module embedding $\phi\colon K[\partial_1, \ldots, \partial_n] \cdot f \to K^r$ we know the vector $\phi(f) \in K^r$ and the companion matrices $A_1, \ldots, A_n \in K^{r\times r}$ describing the action of $\partial_1, \ldots, \partial_n$ on K^r, we can also compute generators of the ideal $\operatorname{ann}(f) \subseteq K[\partial_1, \ldots, \partial_n]$. This works as follows. First observe that for any given term $\partial_1^{e_1} \cdots \partial_n^{e_n}$ we can use $\phi(f)$ as well as the companion matrices $A_1, \ldots, A_n$ to compute the vector $\phi(\partial_1^{e_1} \cdots \partial_n^{e_n} \cdot f) \in K^r$. Therefore, given any finite number of such terms, say $\tau_1, \ldots, \tau_m$, we can decide whether f has an annihilating operator consisting of these terms by checking whether the vectors $\phi(\tau_1 \cdot f), \ldots, \phi(\tau_m \cdot f) \in K^r$ are linearly dependent over K.

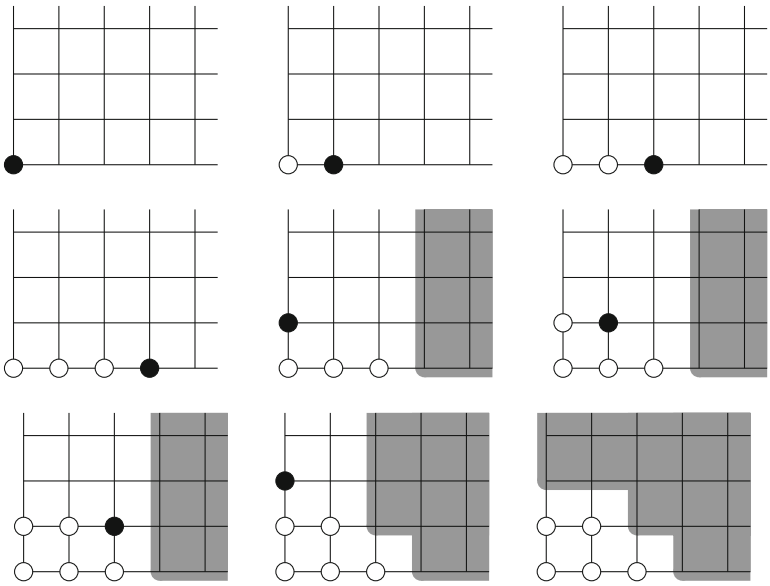
Clearly, for any $p_1, \ldots, p_m \in K$ we have $p_1\tau_1 + \cdots + p_m\tau_m \in \operatorname{ann}(f)$ if and only if $(p_1\tau_1 + \cdots + p_m\tau_m) \cdot f = 0$ if and only if $\phi((p_1\tau_1 + \cdots + p_m\tau_m) \cdot f) = 0$ (since ϕ is supposed to be injective) if and only if $p_1\phi(\tau_1 \cdot f) + \cdots + p_m\phi(\tau_m \cdot f) = 0$.

Secondly, we need a systematic way to choose candidate terms $\tau_1, \ldots, \tau_m$. On the set of all terms $\partial_1^{e_1} \cdots \partial_n^{e_n}$, we define an ordering $\leq$ with the property that $1 = \partial_1^0 \cdots \partial_n^0$ is the smallest element and $\tau_1 \leq \tau_2 \Rightarrow \sigma\tau_1 \leq \sigma\tau_2$ for all terms σ, τ_1, τ_2. Such an order is called a *term order*. With respect to a term order, every nonzero operator $L \in K[\partial_1, \ldots, \partial_n]$ has a maximal term, called the *leading term* of the operator and denoted by $\operatorname{lt}(L)$. By the second defining property of term orders, we have $\operatorname{lt}(\tau L) = \tau \operatorname{lt}(L)$ for every term τ and every nonzero operator $L \in K[\partial_1, \ldots, \partial_n]$. Moreover, it follows from the theory of Gröbner bases (discussed more deeply in the next section) that if I is an ideal, then a basis of the K-vector space $K[\partial_1, \ldots, \partial_n]/I$ is given by the equivalence classes of all terms τ that are not the leading term of an element of I. These observations give rise to the following general procedure for finding generators of an ideal.

Algorithm 4.73 *(FGLM)*
Input: *A term order and a method to determine the set of all K-linear combinations of a given finite set of terms $\tau_1, \ldots, \tau_m$ that belong to I, for a certain ideal $I \subseteq K[\partial_1, \ldots, \partial_n]$.*
Output: *An ideal basis of I and a K-vector space basis of $K[\partial_1, \ldots, \partial_n]/I$.*

1 *Set $B = \emptyset$ and $G = \emptyset$.*
2 *while there exist terms $\partial_1^{e_1} \cdots \partial_n^{e_n}$ which are not in B and not a multiple of some $\operatorname{lt}(g)$ for $g \in G$ do*
3 *Let τ be the smallest such term (with respect to the given term order).*
4 *Search for a nontrivial K-linear combination of $B \cup \{\tau\}$ that belongs to I.*
5 *if there is one then*
6 *Add this relation to G.*
7 *otherwise*
8 *Add τ to B.*
9 *Return G and $\{ [b] : b \in B \}$.*

Example 4.74 Here is a possible trace of the algorithm. We choose the term order defined by $\partial_1^{u_1}\partial_2^{u_2} < \partial_1^{v_1}\partial_2^{v_2}$ if $u_1 < v_1 \vee (u_1 = v_1 \wedge u_2 < v_2)$. In the figures below, a term $\partial_1^{u_1}\partial_2^{u_2}$ corresponds to a point $(u_1, u_2) \in \mathbb{N}^2$. Terms in B are depicted as open circles, the term τ chosen in line 3 of the algorithm is depicted as a filled-in circle, and the terms in the shaded area are multiples of leading terms of elements of G. The nth figure ($n = 1, \ldots, 8$) shows the situation right after the nth execution of line 3. The last figure shows the situation when the while loop has terminated.



□

We have to explain why Algorithm 4.73 terminates and why its output is indeed an ideal basis. For the termination, we need to show that both branches of the if statement in lines 6 and 8 can only be executed a finite number of times. For the statement in line 8, this is easy to see when $K[\partial_1, \ldots, \partial_n]/I$ is a finite dimensional K-vector space, because the equivalence classes of the elements of B are by construction always linearly independent over K. If the quotient $K[\partial_1, \ldots, \partial_n]/I$ has infinite dimension, then the algorithm does not terminate. Regardless of whether the dimension of $K[\partial_1, \ldots, \partial_n]/I$ is finite or not, the statement in line 6 can only be executed a finite number of times, although this is not totally obvious. The reason is the so-called *Dickson's lemma*, which says that every sequence $\tau_1, \tau_2, \tau_3, \ldots$ of terms such that no term τ_i is a multiple of any of its predecessors $\tau_1, \ldots, \tau_{i-1}$ must be finite. Note that this condition applies to the sequence of leading terms $\mathrm{lt}(g)$ added to G by the second part of the termination condition of the while loop.

Theorem 4.75 *Algorithm 4.73 is correct.* □

Proof We have to show that G is a basis of the ideal and that $\{[b] : b \in B\}$ is a vector space basis of the quotient.

It is clear that every element of G belongs to I, because only elements of I are added to G during the algorithm. It is also clear that $\{[b] : b \in B\}$ is linearly independent, because only terms are added to B that do not produce a linear dependence.

In order to show that G generates I and that $\{[b] : b \in B\}$ generates $K[\partial_1, \ldots, \partial_n]/I$, we show that every element L of $K[\partial_1, \ldots, \partial_n]$ is equivalent modulo $\langle G \rangle$ to a linear combination of elements of B. By the linear independence

of B modulo I if and only if this linear combination is zero, this implies that L belongs to I. It also implies that B generates the quotient as a K-vector space, because any element of the quotient which was not in the (sub)space generated by the equivalence classes of elements of B would give rise to a counterexample.

Suppose the contrary, that there are elements of $K[\partial_1, \ldots, \partial_n]$ that are not equivalent modulo $\langle G \rangle$ to a linear combination of elements of B, and among them, let L be one that is minimal in the sense that its largest term τ not contained in B is as small as possible with respect to the term order. By the termination condition of the while loop, any term not contained in B is a multiple of $\mathrm{lt}(g)$ for some $g \in G$. Therefore, there are $p \in K$, a term σ, and a $g \in G$ such that $L - p\sigma g$ does not contain τ. Since L is by assumption not equivalent modulo $\langle G \rangle$ to a linear combination of elements of B, and L is equivalent modulo $\langle G \rangle$ to $L - p\sigma g$, the operator $L - p\sigma g$ must still contain some term that is not in B. However, all such terms must be smaller than $\tau = \sigma\, \mathrm{lt}(g)$, in contradiction to the minimality assumption on L. ∎

Exercises

1. Show that $\dim_{C(x,y)} C(x, y)[D_x, D_y]/\operatorname{ann}(x + \exp(x + y)) = 2$.

2⋆. Show that x^n is D-finite and that $\log(n + x)$ is not.

3. For every $\alpha \in \mathbb{Q}$, determine a basis of the annihilator of $(x^3(x + y)y^2)^\alpha$ with respect to $C(x, y)[D_x, D_y]$.

4. In the proof of Theorem 4.69 we assumed that f belongs to the basis. What if $f = 0$?

5⋆. Show that a sequence $(a_{k_1,\ldots,k_n})_{k_1,\ldots,k_n=0}^{\infty}$ is holonomic with respect to the algebra $C[k_1, \ldots, k_n][S_{k_1}, \ldots, S_{k_n}]$ if and only if its generating function

$$\sum_{k_1,\ldots,k_n=0}^{\infty} a_{k_1,\ldots,k_n} x_1^{k_1} \cdots x_n^{k_n} \in C[[x_1, \ldots, x_n]]$$

is holonomic w.r.t. $C[x_1, \ldots, x_n][D_{x_1}, \ldots, D_{x_n}]$.

Hint: First show that both statements are equivalent to holonomy of the generating function with respect to $C[x_1, \ldots, x_n][\theta_{x_1}, \ldots, \theta_{x_n}]$, where θ_{x_i} acts as the Euler derivation $x_i D_{x_i}$ $(i = 1, \ldots, n)$.

6. The goal of this exercise is to show that a D-finite object depending on n variables can always be viewed as a D-finite object in $n + 1$ variables. To set things

up, consider an Ore algebra $K[\partial_1, \ldots, \partial_n, \partial_{n+1}]$ and let F be a module for the subalgebra $K[\partial_1, \ldots, \partial_n]$.

a. Show that F becomes a $K[\partial_1, \ldots, \partial_n, \partial_{n+1}]$-module by defining $\partial_{n+1} \cdot f = 0$ for every $f \in F$.

b. Show that if $f \in F$ is D-finite with respect to $K[\partial_1, \ldots, \partial_n]$, it is also D-finite with respect to $K[\partial_1, \ldots, \partial_n, \partial_{n+1}]$.

7. Prove or disprove: If a sequence $(a_{n,k})_{n,k=0}^{\infty}$ is such that for every fixed $k \in \mathbb{N}$ the univariate sequence $(a_{n,k})_{n=0}^{\infty}$ is D-finite, then $(a_{n,k})_{n,k=0}^{\infty}$ is D-finite as a bivariate sequence.

8⋆. **a.** Let $r_1, \ldots, r_n \in C(x)$ be such that their denominators are pairwise coprime. Show that $r_1, \ldots, r_n$ are linearly independent over C.

b. Let $m \in \mathbb{Q} \setminus \mathbb{Z}$. Show that the polynomials $(m+u)^2 + (x+v)^2$ for $u, v \in \mathbb{Z}$ are pairwise coprime.

9. Prove or disprove: $1/(n^2 - k^2)$ is holonomic.

10⋆. Prove or disprove: $1/(nk+1)$ is holonomic.

11. Prove or disprove: If $I, J \subseteq C[x_1, \ldots, x_n, \partial_{x_1}, \ldots, \partial_{x_n}]$ are nontrivial (i.e., different from $\langle 1 \rangle$) and holonomic, then $I \subseteq J$ implies $I = J$.

12⋆⋆⋆. Show that $\sum_{k=1}^{n} \frac{1}{x+k}$ is not D-finite.

13. Show that the Legendre polynomials are holonomic in n and x. *Hint:* Guess and prove appropriate annihilating operators.

14. Let the linear functions $g_1, \ldots, g_n \colon \mathbb{Q}^m \to \mathbb{Q}$ be linearly independent over C. Show that $g_1, \ldots, g_n$ are algebraically independent.

15⋆. We have seen that if a multivariate sequence is D-finite, its generating function need not be D-finite. How about the converse: If a multivariate power series is D-finite, does its coefficient sequence have to be D-finite?

16⋆⋆. Let F be a $C(x, y)[D_x, D_y]$-module and $f \in F$.

a. Show that $\operatorname{ann}(f) = \langle D_x - x, D_y - x \rangle$ implies $f = 0$.

b. Suppose that there is an embedding $\phi \colon C(x, y)[D_x, D_y] \cdot f \to C(x, y)^2$ with the companion matrices $A_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $A_y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Show that $f = 0$.

17⋆. Let $A = K[\partial_1, \ldots, \partial_n]$ be an Ore algebra, F be an A-module and let $m \colon F \times F \to F$ be a bilinear map. Let $f, g \in F$ be D-finite. Suppose there is an embedding $\phi \colon A \cdot f \to K^2$ with $\phi(f) = (1, 0) \in K^2$ and companion matrices $A_1, \ldots, A_n \in K^{2\times 2}$ and an embedding $\psi \colon A \cdot g \to K^2$ with $\psi(g) = (1, 0) \in K^2$ and companion matrices $B_1, \ldots, B_n \in K^{2\times 2}$. We want to construct an embedding $\chi \colon A \cdot m(f, g) \to K^4$ with $\chi(m(f, g)) = (1, 0, 0, 0)$. Find a companion matrix for the action of ∂_i

a. if $\partial_i \cdot m(a, b) = m(\partial_i \cdot a, \partial_i \cdot b)$ for all $a, b \in F$,

b. if $\partial_i \cdot m(a, b) = m(\partial_i \cdot a, b) + m(a, \partial_i \cdot b)$ for all $a, b \in F$.

18. Suppose that $f \colon \mathbb{Z}^2 \to C[[x, y]]$ is D-finite. Show that $f(3n + 5k, 2n - k, \sqrt{1 - xy}, x^2 + y^2)$ is D-finite as well.

19⋆. Let $u, v \in \mathbb{Q}^n$ and define $f\colon \mathbb{Z}^n \to C$ by

$$f(k_1, \ldots, k_n) = \begin{cases} 1 \text{ if } \big((k_1, \ldots, k_n) - u\big) \cdot v \geq 0, \\ 0 \text{ otherwise,} \end{cases}$$

so that f is 1 for the points in a certain halfspace defined by u and v, and 0 outside of this halfspace. Show that f is holonomic.

20. Let $I = \langle (1 - xy)D_x + y^2 D_y - y, (1 - xy)D_y^2 + x^2 y D_y - x^2 \rangle \subseteq C(x, y)[D_x, D_y]$. Construct an embedding $\phi\colon C(x, y)[D_x, D_y]/I \to C(x, y)^2$.

21. For a certain D-finite sequence f we have an embedding $\phi\colon C(n, k)[S_n, S_k] \cdot f \to C(n, k)^2$ with $\phi(f) = \binom{1}{1}$ and the companion matrices

$$A_n = \begin{pmatrix} \frac{2(n+1)(2n+1)}{(n+k+1)(n-k+1)} & 0 \\ 0 & 1 \end{pmatrix}, \quad A_k = \begin{pmatrix} \frac{n-k}{n+k+1} & 0 \\ 0 & 1 \end{pmatrix}$$

describing the action of S_n, S_k, respectively. Compute a basis of an ideal $I \subseteq C(n, k)[S_n, S_k]$ with $\dim_{C(n,k)} C(n, k)[S_n, S_k]/I = 2$ and $I \subseteq \operatorname{ann}(f)$. Can we tell whether $I = \operatorname{ann}(f)$?

22⋆. For the Legendre polynomials $P_n(x)$ we have $\operatorname{ann}(P_n(x)) = \langle (n + 1)S_n - (x^2 - 1)D_x - (n+1)x, (x^2 - 1)D_x^2 + 2x D_x - n(n+1) \rangle \subseteq C(n, x)[S_n, D_x]$. Compute an ideal basis of $\operatorname{ann}(P_n(x)^2)$.

23. Let $a(x, y)$ be an algebraic function satisfying the polynomial equation $a(x, y)^4 - x a(x, y)^2 + y = 0$. Find an ideal $I \subseteq C(x, y)[D_x, D_y]$ with $\operatorname{ann} a(x, y) \subseteq I$ and $\dim_{C(x,y)} C(x, y)[D_x, D_y]/I = 2$.

24. Show that the sum of two holonomic sequences is holonomic.

25⋆. Let $a(x, y) \in C[[x, y]]$ be D-finite and such that $a(x, y) = a(y, x)$.

a. Show that there are nonzero annihilating operators of $a(x, y)$ which are invariant under exchanging x with y and D_x with D_y.

b. Show that not all annihilating operators of $a(x, y)$ have this property.

26⋆. (Shaoshi Chen)

a. If $a \in C[[t]]$ is such that $a(xy) \in C[[x, y]]$ is D-finite, then a is D-finite in the sense of Chap. 3.

b. If $a \in C[[x, y]]$ is such that $(x D_x - y D_y)^s \cdot a = 0$ for some $s \in \mathbb{N}$, then there exists $b \in C[[t]]$ such that $a(x, y) = b(xy)$.

27⋆⋆. Show that there is no algorithm which for any given

$$L \in C[x_1, \ldots, x_n][D_{x_1}, \ldots, D_{x_n}]$$

decides whether there exists a nonzero polynomial $p \in C[x_1, \ldots, x_n]$ such that $L \cdot p = 0$.

Hint: You may use Matiyasevich's theorem, which says that there is no algorithm which for any given polynomial $p \in \mathbb{Z}[x_1, \ldots, x_n]$ decides whether there is a tuple $(\xi_1, \ldots, \xi_n) \in \mathbb{N}^n$ such that $p(\xi_1, \ldots, \xi_n) = 0$.

References

Multivariate D-finite functions were first considered by Zeilberger [466]. He calls them multi-D-finite in the differential case and multi-P-recursive in the shift case. Ore algebras were first used to describe multivariate objects by Chyzak and Salvy [157]. They propose the notion *∂-finite* for what we call D-finite in Definition 4.67. Lipshitz [314] observed that generating functions of D-finite sequences need not be D-finite, which led Zeilberger to use holonomy instead of D-finiteness in his paper [468].

Holonomy was introduced by Bernstein [58] and has developed into a rather sophisticated theory [69, 166, 260, 377], of which we hardly make any use here. A fundamental result of the theory, known as Bernstein's inequality, is that for a proper left ideal $I \subsetneq C[x_1, \ldots, x_n][D_{x_1}, \ldots, D_{x_n}]$, it cannot happen that $I \cap C[U] \neq \{0\}$ for every subset $U \subseteq \{x_1, \ldots, x_n, D_{x_1}, \ldots, D_{x_n}\}$ with $|U| = n + 2$. Algebraically speaking, this means that the dimension of any proper left ideal of $C[x_1, \ldots, x_n][D_{x_1}, \ldots, D_{x_n}]$ is at least n. In view of this result, a proper ideal is holonomic if it has smallest possible dimension. Another peculiar fact due to Stafford [307, 411] is that every left ideal of $C[x_1, \ldots, x_n][D_{x_1}, \ldots, D_{x_n}]$ can be generated by only two elements.

We have seen in Sect. 2.2 that the dimension of the solution space of a linear recurrence with polynomial coefficients may exceed the order of the equation. The same is true in the multivariate case. Abramov and Petkovšek show that the solution space of a system of recurrences of first order can have any dimension [19].

Algorithm 4.73 is named after the initials of Faugère, Gianni, Lazard, and Mora, who propose this technique in [191] for changing the term order of a Gröbner basis. Essentially the same idea was already used a decade earlier by Buchberger and Möller for constructing Gröbner bases of ideals with finitely many solutions [118]. Dickson's lemma comes from [171].

4.6 Gröbner Bases

The annihilating operators of a D-finite object form a left ideal of the operator algebra. In the univariate case, the operator algebra is typically an Ore algebra of the form $K[\partial]$, which is a principal left ideal domain, so every ideal can be described by a single generator. It consists of all the left multiples of such a generator. In the case of several variables, an ideal is not necessarily generated by a single operator, but it remains true that every ideal can be described by a finite set of

generators. This is Hilbert's basis theorem, which was originally formulated for commutative polynomial rings $C[x_1, \ldots, x_n]$ over a field but also holds for Ore algebras $K[\partial_1, \ldots, \partial_n]$.

If we are given a finite set $B \subseteq K[\partial_1, \ldots, \partial_n]$, it is not necessarily easy to answer questions about the left ideal $\langle B \rangle$ it generates. Even the answer to the question of whether or not the ideal contains 1 may not be obvious at first glance. Gröbner bases theory gives a finite set G of generators for an ideal $I \subseteq K[\partial_1, \ldots, \partial_n]$ such that many questions about I can be easily answered by looking at G. The theory is rich and has many applications, especially in the commutative case, for which it was first developed. There are several excellent textbooks exclusively devoted to Gröbner bases for commutative polynomial rings, and since the theory extends almost literally to the case of Ore algebras, we only give a minimal discussion here.

We are primarily interested in two kinds of noncommutative polynomial rings: Ore algebras $C(x_1, \ldots, x_m)[\partial_1, \ldots, \partial_k]$ in which $x_1, \ldots, x_m$ belong to the ground field, and Ore algebras $C[x_1, \ldots, x_m][\partial_1, \ldots, \partial_k]$ in which $x_1, \ldots, x_m$ are also considered as polynomial variables. In order to cover both cases with a common notation, let us write $K[X_1, \ldots, X_n]$ for the rings under consideration, where K may refer to either $C(x_1, \ldots, x_m)$ or C, and the variables $X_1, \ldots, X_n$ may refer either just to $\partial_1, \ldots, \partial_k$ or to $x_1, \ldots, x_m, \partial_1, \ldots, \partial_k$.

It remains true that every element of $K[X_1, \ldots, X_n]$ can be written as a (left-) K-linear combination of terms of the form $X_1^{e_1} \cdots X_n^{e_n}$ with $e_1, \ldots, e_n \in \mathbb{N}$, but we may no longer assume that the product of two terms is again a term. For example, in $C[x, y][D_x, D_y]$ we have $D_x x = x D_x + 1$. We therefore need to refine the definition of term orders used in the previous section. Like before, if $\leq$ is a total order on the set of all terms, we call the largest term with respect to $\leq$ appearing in an element $p \in K[X_1, \ldots, X_n] \setminus \{0\}$ the *leading term* of p and denote it by $\mathrm{lt}(p)$. We also define the *leading coefficient* $\mathrm{lc}(p) := [\mathrm{lt}(p)]p$ and the *leading monomial* $\mathrm{lm}(p) := \mathrm{lc}(p)\,\mathrm{lt}(p)$ of p. The *leading exponent* $\mathrm{lexp}(p)$ of p is defined as the vector $(e_1, \ldots, e_n)$ such that $\mathrm{lt}(p) = X_1^{e_1} \cdots X_n^{e_n}$. Note that all of these notions depend on the choice of the order $\leq$.

A total order $\leq$ on the set of terms is now called a *term order* (or *monomial order* or *admissible order*) if (i) $1 = X_1^0 \cdots X_n^0$ is the minimal element with respect to $\leq$; (ii) $\tau \leq \sigma \Rightarrow \mathrm{lt}(\rho\tau) \leq \mathrm{lt}(\rho\sigma)$ for all terms τ, σ, ρ; (iii) for all i, j with $i < j$ there exist $u \in K \setminus \{0\}$ and $v \in K[X_1, \ldots, X_n]$ with $v = 0$ or $\mathrm{lt}(v) < X_i X_j$ such that $X_j X_i = u X_i X_j + v$. This definition differs from the commutative case, where condition (iii) is not needed because it follows from (ii), and where on the right hand side of the implication in (ii) it suffices to say $\rho\tau \leq \rho\sigma$ because products of terms are terms. The adjustments are made in such a way that the rest of the theory carries over seamlessly to the present setting.

We assume from now on that the ring $K[X_1, \ldots, X_n]$ is endowed with a certain fixed term order $\leq$. Once a term order is fixed, we can perform division with remainder. In the commutative case, the algorithm for division with remainder is based on divisibility properties of leading terms. In $K[X_1, \ldots, X_n]$, we cannot easily say that one term is a divisor of another term because products of terms

need not be terms. But we can talk about exponent vectors instead. For two vectors $(e_1, \ldots, e_n), (e'_1, \ldots, e'_n) \in \mathbb{N}^n$, we write $(e_1, \ldots, e_n) \leq (e'_1, \ldots, e'_n)$ if $\forall\, i : e_i \leq e'_i$. Then the divisibility $\mathrm{lt}(p) \mid \mathrm{lt}(q)$ used in the commutative case translates into the relation $\mathrm{lexp}(p) \leq \mathrm{lexp}(q)$. With this notation, the division algorithm can be formulated as follows.

Algorithm 4.76 *(Reduction)*
Input: $p \in K[X_1, \ldots, X_n]$, $G \subseteq K[X_1, \ldots, X_n]$, *a term order* $\leq$.
Output: $r \in K[X_1, \ldots, X_n]$ *such that* $p - r \in \langle G \rangle$ *and* r *contains no terms* τ *with* $\mathrm{lexp}(g) \leq \mathrm{lexp}(\tau)$ *for some* $g \in G$.

1 *Set* $r = 0$.
2 *while* $p \neq 0$ *do*
3 *if there is a* $g \in G$ *and a term* τ *in* p *such that* $\mathrm{lexp}(g) \leq \mathrm{lexp}(\tau)$ *then*
4 *Let* $g \in G$ *and* $\sigma = X_1^{e_1} \cdots X_n^{e_n}$ *be such that* $\mathrm{lt}(\sigma\, \mathrm{lc}(g)^{-1} g) = \tau$.
5 *Set* $p = p - ([\tau]p)\sigma\, \mathrm{lc}(g)^{-1} g$.
6 *else*
7 *Set* $p = p - ([\tau]p)\tau$ *and* $r = r + ([\tau]p)\tau$.
8 *Return* r.

It is easy to see that the algorithm is correct. Indeed, we obviously have $p - r \in \langle G \rangle$ in the beginning, and the property is preserved in every iteration of the loop, regardless of whether line 5 or line 7 is executed. So $p - r \in \langle G \rangle$ is true in the end. Moreoever, in line 7 we only introduce monomials into r whose exponent vectors are not above the leading exponent of any element of G, so this is also true at the end.

It is less clear that the algorithm terminates. Lines 5 and 7 are designed to cancel a term from p, but line 5 may introduce many other terms, which will be smaller in term order than the eliminated term. From a hypothetical infinite run of the algorithm, it can be deduced that there is an infinite strictly descending sequence of terms. Since such a sequence does not exist by Dickson's lemma, the algorithm terminates.

It is also not clear why the output of the algorithm is independent of the choice of g made in line 4, if there are several options. In fact, no claim is made that the output is unique, and in general, different choices of g in line 4 do lead to different output. A starting point of the theory of Gröbner bases is the desire to make the output unique by imposing appropriate restrictions on G. In view of this goal, let us use the notation $\mathrm{red}(p, G)$ for any possible outcome of Algorithm 4.76 when applied to $p \in K[X_1, \ldots, X_n]$ and $G \subseteq K[X_1, \ldots, X_n]$. Note that in view of the non-uniqueness, $\mathrm{red}(\cdot, G)$ is not a function, and $\mathrm{red}(p, G) = r_1$ and $\mathrm{red}(p, G) = r_2$ does not imply $r_1 = r_2$. This is similar to the big-O notation. If the term order is not clear from context, we write $\mathrm{red}_{\leq}(p, G)$ instead of $\mathrm{red}(p, G)$.

The definition of Gröbner bases is motivated by the following theorem.

Theorem 4.77 *Let* $G \subseteq K[X_1, \ldots, X_n]$. *Then the following are equivalent:*

1. *For all* $p \in K[X_1, \ldots, X_n]$ *there exists exactly one* $r \in K[X_1, \ldots, X_n]$ *with* $\operatorname{red}(p, G) = r$.
2. *For all* $p \in \langle G \rangle$ *we have* $\operatorname{red}(p, G) = 0$.
3. *For all* $p \in \langle G \rangle \setminus \{0\}$ *there exists a* $g \in G$ *with* $\operatorname{lexp}(g) \leq \operatorname{lexp}(p)$.
4. *The set of all equivalence classes of terms* $\tau = X_1^{e_1} \cdots X_n^{e_n}$ *with* $\operatorname{red}(\tau, G) = \tau$ *forms a* K*-vector space basis of* $K[X_1, \ldots, X_n]/\langle G \rangle$. □

Proof 1. $\Rightarrow$ 2.: Let $p \in \langle G \rangle$ and let $r = \operatorname{red}(p, G)$. We have to show that $r = 0$.

First observe that for every $u \in K[X_1, \ldots, X_n]$, every $c \in K$, every term $\tau = X_1^{e_1} \cdots X_n^{e_n}$, and every $g \in G$ we have $\operatorname{red}(u + c\tau \operatorname{lc}(g)^{-1} g, G) = \operatorname{red}(u, G)$. This is a consequence of assuming 1. (See Exercise 7 for a detailed argument.)

Since $p \in \langle G \rangle$, there is a way to write $p = \sum_{i=1}^{m} c_i \tau_i \operatorname{lc}(g_i)^{-1} g_i$ with certain ground field elements $c_i \in K$, certain terms τ_i, and certain elements g_i of G (not necessarily pairwise distinct). Applying the observation m times, we find that $0 = \operatorname{red}(0, G) = \operatorname{red}(p - \sum_{i=1}^{m} c_i \tau_i \operatorname{lc}(g_i)^{-1} g_i, G) = r$, as required.

2. $\Rightarrow$ 3.: If $p \neq 0$ reduces to zero, the reduction process must have at least one step. The reduction process as formulated in Algorithm 4.76 is not forced to start with eliminating the leading term of p, but as long as it keeps eliminating smaller terms, the leading term of p will remain unchanged. In order to eventually reach zero, it must at some point choose a g with $\operatorname{lexp}(g) \leq \operatorname{lexp}(p)$ in order to also eliminate the leading term. Therefore, such a g must exist.

3. $\Rightarrow$ 4.: Let B be the set of equivalence classes defined in the statement. It is clear that B generates $K[X_1, \ldots, X_n]/\langle G \rangle$ because for every $p \in K[X_1, \ldots, X_n]$ we have $[p] = [\operatorname{red}(p, G)]$ and $\operatorname{red}(p, G)$ only contains terms τ with $\operatorname{red}(\tau, G) = \tau$, so $[p]$ is a K-linear combination of elements of B. The set B is also K-linearly independent, for if $[p]$ is a K-linear combination of elements of B, we may assume that p is a K-linear combination of terms τ with $\operatorname{red}(\tau, G) = \tau$. The class $[p]$ is zero if and only if $p \in \langle G \rangle$, which by assumption implies $\operatorname{red}(p, G) = 0$. But since p does not contain any terms that can be reduced, it cannot contain any terms at all. From $p = 0$ follows the linear independence of B.

4. $\Rightarrow$ 1.: Let $p \in K[X_1, \ldots, X_n]$ and let $r_1, r_2 \in K[X_1, \ldots, X_n]$ be such that $\operatorname{red}(p, G) = r_1$ and $\operatorname{red}(p, G) = r_2$. Then r_1 and r_2 contain only terms τ with $\operatorname{red}(\tau, G) = \tau$. Moreover, $r_1 - r_2$ is an element of the ideal, so $[r_1 - r_2] = 0$. Since the set B in statement 4 is linearly independent by assumption, it follows that $r_1 = r_2$. ■

Definition 4.78 A set $G \subseteq K[X_1, \ldots, X_n]$ is called a *Gröbner basis* (of the left ideal $\langle G \rangle$) if it satisfies any of the equivalent conditions of Theorem 4.77. □

Every left ideal I of $K[X_1, \ldots, X_n]$ is a Gröbner basis (of itself). The reason is that every $p \in I$ can be reduced to zero in one step. It is also not hard to see that for every left ideal I of $K[X_1, \ldots, X_n]$, there exists a finite Gröbner basis G with $\langle G \rangle = I$. The reason is Dickson's lemma: start with an arbitrary element $g_1 \in I \setminus \{0\}$, then, if possible, choose an element $g_2 \in I \setminus \{0\}$ with $\operatorname{red}(\operatorname{lt}(g_2), \{g_1\}) = \operatorname{lt}(g_2)$, then, if possible, an element $g_3 \in I$ with $\operatorname{red}(\operatorname{lt}(g_3), \{g_1, g_2\}) = \operatorname{lt}(g_3)$, and so on. Because of $\operatorname{lexp}(g_1) > \operatorname{lexp}(g_2) > \operatorname{lexp}(g_3) > \cdots$, the process must come to an

end after finitely many steps. The resulting set $\{g_1, \ldots, g_k\}$ is the desired Gröbner basis.

This argument is not constructive. If we want to compute a Gröbner basis for a given ideal I, we first have to agree what it means for I to be "given". One situation is that we know some finite set $B \subseteq K[X_1, \ldots, X_n]$ such that $I = \langle B \rangle$ and want to know a Gröbner basis G with $I = \langle G \rangle$. This problem is solved by Buchberger's algorithm, which is explained below. Another typical situation is if we can apply Algorithm 4.73 with our knowledge about I. Observe that the ideal basis returned by Algorithm 4.73 is always a Gröbner basis, which can be seen using characterization 4 of Theorem 4.77.

Knowing a Gröbner basis, we can also meet the input requirements of Algorithm 4.73. Namely, in order to find K-linear combinations modulo $\langle G \rangle$ among some given terms $\tau_1, \ldots, \tau_m$, we can make an ansatz $u_1\tau_1 + \cdots + u_m\tau_m$ for undetermined $u_i \in K$ and force $\operatorname{red}(u_1\tau_1 + \cdots + u_m\tau_m, G) = u_1 \operatorname{red}(\tau_1, G) + \cdots + u_m \operatorname{red}(\tau_m, G) = 0$ by equating coefficients of like terms. But why would we want to compute a Gröbner basis if we already have one? One reason could be that we only know a Gröbner basis with respect to a certain term order $\leq_1$ but we would need a Gröbner basis for the same ideal with respect to some other term order $\leq_2$. Another situation is when we know a Gröbner basis for some ideal(s) but would like to compute a (Gröbner) basis for a different ideal. For example, suppose we already know Gröbner bases G_1, G_2 of $\operatorname{ann}(f_1)$ and $\operatorname{ann}(f_2)$ for two D-finite functions f_1 and f_2. Then $\operatorname{ann}(f_1) \cap \operatorname{ann}(f_2)$ is an ideal of annihilating operators for $f_1 + f_2$, and we can use Algorithm 4.73 to compute generators for it. As input to the algorithm, we can use a procedure which for given terms $\tau_1, \ldots, \tau_m$ computes $u_1, \ldots, u_m \in K$ such that $u_1 \operatorname{red}(\tau_1, G_1) + \cdots + u_m \operatorname{red}(\tau_m, G_1) = 0$ and $u_1 \operatorname{red}(\tau_1, G_2) + \cdots + u_m \operatorname{red}(\tau_m, G_2) = 0$. Other closure properties can be executed in a similar fashion.

Of particular interest in the context of D-finite functions is characterization 4 of Theorem 4.77. Suppose we have a function f for which we know a Gröbner basis G of the ideal $\operatorname{ann}(f) \subseteq K[\partial_1, \ldots, \partial_n]$. By Definition 4.63, f is D-finite if and only if $\dim_K K[\partial_1, \ldots, \partial_n]/\operatorname{ann}(f) < \infty$. According to Exercise 9, this is the case if and only if for every $i \in \{1, \ldots, n\}$ there exists a $g \in G$ whose leading term is a power of ∂_i, a condition that can easily be checked by inspection. Gröbner bases can also be used for checking whether an ideal is holonomic. Suppose we know a basis B of the ideal $I \subseteq C[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$. According to Definition 4.67, I is holonomic if and only if $I \cap C[U] \neq \{0\}$ for every $U \subseteq \{x_1, \ldots, x_n, \partial_1, \ldots, \partial_n\}$ with $|U| = n + 1$. For each such U, select a term order for which terms consisting only of variables from U are smaller than terms containing other variables. Such a term order is called an *elimination order* (for U). It follows easily from the defining properties of Theorem 4.77 that the ideal $I \cap C[U]$ is generated by $G \cap C[U]$ whenever G is a Gröbner basis with respect to an elimination order for U. In particular, $I \cap C[U] \neq \{0\}$ if and only if $G \cap C[U] \neq \emptyset$.

For computing a Gröbner basis from an arbitrary (but finite) given ideal basis, none of the conditions of Theorem 4.77 are particularly useful, because all are statements about infinitely many cases that cannot simply be checked one by one.

Buchberger's algorithm is based on a different characterization which only affects finitely many cases. In the commutative case, the *S-polynomial* of two polynomials $p, q \in C[x_1, \ldots, x_n] \setminus \{0\}$ is defined as

$$\operatorname{spol}(p, q) = \frac{\operatorname{lcm}(\operatorname{lt}(p), \operatorname{lt}(q))}{\operatorname{lm}(p)} p - \frac{\operatorname{lcm}(\operatorname{lt}(p), \operatorname{lt}(q))}{\operatorname{lm}(q)} q.$$

In a sense, this is the smallest possible linear combination of p and q which induces a cancellation of the leading monomials of p and q. Note that both summands on the right have the leading term $\operatorname{lcm}(\operatorname{lt}(p), \operatorname{lt}(q))$. The same idea is used in the non-commutative case, but we should adapt the notation a bit, because speaking of the "least common multiple" of terms does not seem appropriate if the product of two terms is not necessarily again a term. It is safer to talk about exponent vectors. For two vectors $u = (u_1, \ldots, u_n)$, $v = (v_1, \ldots, v_n) \in \mathbb{N}^n$, define $\max(u, v) = (\max(u_1, v_1), \ldots, \max(u_n, v_n))$ and write X^u for $X_1^{u_1} \cdots X_n^{u_n}$. Then the *S-polynomial* of $p, q \in K[X_1, \ldots, X_n] \setminus \{0\}$ is defined as

$$\begin{aligned} \operatorname{spol}(p, q) = {} & X^{\max(\operatorname{lexp}(p), \operatorname{lexp}(q)) - \operatorname{lexp}(p)} \operatorname{lc}(p)^{-1} p \\ & - X^{\max(\operatorname{lexp}(p), \operatorname{lexp}(q)) - \operatorname{lexp}(q)} \operatorname{lc}(q)^{-1} q. \end{aligned}$$

Again, this is the smallest possible way to let the leading monomials of p and q cancel. With this definition of S-polynomials, Buchberger's characterization of Gröbner bases and his algorithm for computing Gröbner bases carry over literally from the commutative case.

Theorem 4.79 *A set $G \subseteq K[X_1, \ldots, X_n]$ is a Gröbner basis if and only if* $\operatorname{red}(\operatorname{spol}(p, q), G) = 0$ *for all $p, q \in G$.* □

Proof The direction "⇒" follows directly from Definition 4.78. To show "⇐", let $p \in \langle G \rangle$ be such that $\operatorname{red}(p, G) = p$, i.e., no term appearing in p can be matched with the leading term of a multiple of an element of G. We show that $p = 0$.

Since $p \in \langle G \rangle$, there are $g_1, \ldots, g_m \in G$ and $p_1, \ldots, p_m \in K[X_1, \ldots, X_n]$ such that $p = p_1 g_1 + \cdots + p_m g_m$. Since p cannot be reduced, the leading terms $\operatorname{lt}(p_i g_i)$ do not occur in p, so there must be some cancellation on the right hand side. We show that in fact the entire right hand side cancels.

Suppose otherwise. Let $\tau_i = \operatorname{lt}(p_i g_i)$ for $i = 1, \ldots, m$ and assume without loss of generality that the indexing is such that $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_m$. We may further assume, also without loss of generality, that the coefficients $p_1, \ldots, p_m$ are chosen in such a way that τ_1 is as small as possible.

In view of the required cancellation, we must have $\tau_1 = \tau_2 = \cdots = \tau_k > \tau_{k+1}$ for some $k \geq 2$. We may further assume, again without loss of generality, that among all possible choices $p_1, \ldots, p_m$ that yield the minimal τ_1, our choice is made such that k is minimal.

We have

$$\operatorname{lm}(p_k)g_k = \operatorname{lm}(p_k)\operatorname{lc}(g_k)\operatorname{lc}(g_k)^{-1}g_k = (u\operatorname{lt}(p_k) + v)\operatorname{lc}(g_k)^{-1}g_k$$

for some $u \in K$ and some $v \in K[X_1, \ldots, X_n]$ with $v = 0$ or $\operatorname{lt}(v) < \operatorname{lt}(p_k)$. There are also terms σ, μ such that

$$\operatorname{lt}(p_k)\operatorname{lc}(g_k)^{-1}g_k - \sigma\operatorname{lc}(g_{k-1})^{-1}g_{k-1} = \mu\operatorname{spol}(g_k, g_{k-1}).$$

Since $\operatorname{red}(\operatorname{spol}(g_k, g_{k-1}), G) = 0$ by assumption, there exist polynomials $q_1, \ldots, q_m$ with

$$u\mu\operatorname{spol}(g_k, g_{k-1}) = q_1g_1 + \cdots + q_mg_m$$

and $\operatorname{lt}(q_ig_i) < \mu X^{\max(\operatorname{lexp}(g_k),\operatorname{lexp}(g_{k-1}))} = \tau_1$ for $i = 1, \ldots, m$. Using $\operatorname{lm}(p_k) - v\operatorname{lc}(g_k)^{-1} = u\operatorname{lt}(p_k)\operatorname{lt}(g_k)^{-1}$, we get

$$\begin{aligned}
p = {} & \big(p_1 \qquad\qquad\qquad\qquad + q_1 \quad\big)g_1 \\
& \quad\vdots \\
& + \big(p_{k-2} \qquad\qquad\qquad\qquad + q_{k-2}\big)g_{k-2} \\
& + \big(p_{k-1} + u\sigma\operatorname{lc}(g_{k-1})^{-1} \qquad + q_{k-1}\big)g_{k-1} \\
& + \big(p_k \quad - (\operatorname{lm}(p_k) - v\operatorname{lc}(g_k)^{-1}) + q_k \quad\big)g_k \\
& + \big(p_{k+1} \qquad\qquad\qquad\qquad + q_{k+1}\big)g_{k+1} \\
& \quad\vdots \\
& + \big(p_m \underbrace{\qquad\qquad\qquad\qquad + q_m \quad\big)}_{=0}g_m.
\end{aligned}$$

This new representation of p violates the minimality assumption on k, or, if $k = 2$, the minimality assumption of τ_1. ■

Algorithm 4.80 *(Buchberger)*
Input: *A finite set* $B \subseteq K[X_1, \ldots, X_n] \setminus \{0\}$ *and a term order* $\leq$.
Output: *A finite Gröbner basis* $G \subseteq K[X_1, \ldots, X_n]$ *with respect to* $\leq$ *such that* $\langle B\rangle = \langle G\rangle$.

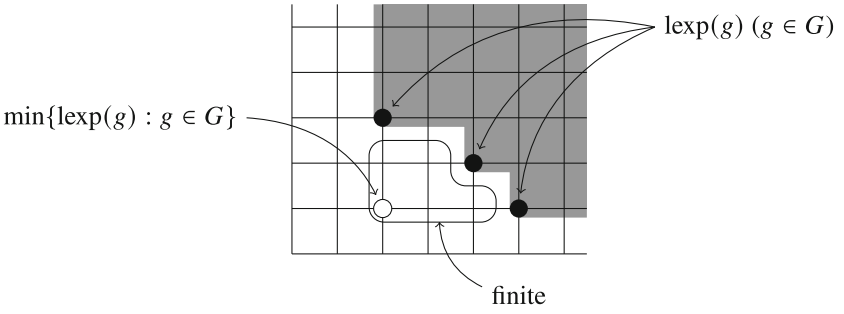
Set $G = B = \{b_1, \ldots, b_m\}$.
Set $P = \{(b_i, b_j) : 1 \leq i < j \leq m\}$.
while $P \neq \emptyset$ *do*
  Choose a pair $(p, q) \in P$ *and set* $P = P \setminus \{(p, q)\}$.
  Compute $h = \operatorname{red}(\operatorname{spol}(p, q), G)$.
  if $h \neq 0$ *then*

7 *Set* $P = P \cup \{(p, h) : p \in G\}$ *and* $G = G \cup \{h\}$.
8 *Return* G.

Theorem 4.81 *Algorithm 4.80 is correct and terminates.* □

Proof Correctness follows from Theorem 4.79, because the algorithm terminates if G is such that all S-polynomials reduce to zero. Whenever the algorithm encounters an S-polynomial that does not reduce to zero, it adds the remainder h to G, so that h can then be reduced further to zero in one step. Recall that the notation $\mathrm{red}(\cdot, G)$ refers to any possible output of the reduction procedure and is not unique as long as G is not a Gröbner basis. However, if there is some way to reduce a certain S-polynomial $\mathrm{spol}(p, q)$ to zero at a given stage of the algorithm, then this feature is not harmed if we add further elements to G later during the computation. Therefore, when G gets updated, it is not necessary to reconsider the S-polynomials that have been handled up to that point.

For the termination, it suffices to show that line 7 cannot be executed infinitely often. At every execution of line 7, consider the vector $\min\{\mathrm{lexp}(g) : g \in G\} \in \mathbb{N}^n$. Since h cannot be reduced any further by G, we have $\mathrm{lexp}(g) \not\leq \mathrm{lexp}(h)$ for all $g \in G$. This does not necessarily mean that adding h to G leads to a drop in one of the coordinates of $\min\{\mathrm{lexp}(g) : g \in G\}$, but since there are only finitely many points above $\min\{\mathrm{lexp}(g) : g \in G\}$ and below the $\mathrm{lexp}(g)$ $(g \in G)$, the vector can cease to change only finitely many times. Hence, at least every now and then during a long execution, we must observe that at least one coordinate of $\min\{\mathrm{lexp}(g) : g \in G\}$ strictly decreases. Since no coordinate of the vector can ever increase, we must reach $(0, \ldots, 0)$ after finitely many steps, unless the algorithm terminates before. At this point, it terminates in any case. ■



Example 4.82 Consider the ideal $I = \langle b_1, b_2, b_3 \rangle \subseteq C(n, k)[S_n, S_k]$ with

$$b_1 = S_n^2 S_k + S_n S_k^2 - 3S_n S_k - S_n - S_k^2 + S_k + 2,$$

$$b_2 = 5(k - n - 2)S_n^2 - (2n - 7)S_n S_k - (9k - 14n - 19)S_n + (k + 2)S_k^2 + (2k + n - 5)S_k + k - 8n - 16,$$

$$b_3 = (5k - 3n - 2)S_n S_k - (k - n - 1)S_n - (k + 2)S_k^2 \\ - (2k - 4n - 5)S_k - k - 2n + 1.$$

We want to compute a Gröbner basis with respect to the term order $\leq$ defined by $S_n^a S_k^b \leq S_n^u S_k^v$ if $a < u$ or $a = u$ and $b \leq v$. The terms in b_1, b_2, b_3 are already sorted according to this order.

We set $P = \{(b_1, b_2), (b_1, b_3), (b_2, b_3)\}$ and select (b_1, b_2) as the first pair. Its S-polynomial has the form

$$b_1 - S_k \tfrac{1}{5(k-n-2)} b_2 = (\ldots)S_n^2 + (\ldots)S_n S_k^2 + (\ldots)S_n S_k + (\ldots)S_n \\ + (\ldots)S_k^3 + (\ldots)S_k^2 + (\ldots)S_k + (\ldots),$$

where the $\ldots$ are certain elements of $C(n, k)$ that we suppress here. Using Algorithm 4.76, the S-polynomial can be reduced to an operator of the form

$$(\ldots)S_n + (\ldots)S_k^3 + (\ldots)S_k^2 + (\ldots)S_k + (\ldots).$$

As it is nonzero, we call it b_4 and add it to the basis. We also add the pairs (b_1, b_4), (b_2, b_4), (b_3, b_4) to P.

For the next pair, (b_1, b_3), the reduction of the S-polynomial produces an operator of the form

$$(\ldots)S_k^3 + (\ldots)S_k^2 + (\ldots)S_k + (\ldots),$$

and since it is nonzero, we call it b_5 and add it to the basis. We also add the pairs (b_1, b_5), (b_2, b_5), (b_3, b_5), (b_4, b_5) to P.

For the next pair, (b_2, b_3), the S-polynomial reduces to zero, so we get nothing new.

Next, (b_1, b_4) has an S-polynomial which reduces to an operator of the form

$$(\ldots)S_k^2 + (\cdots)S_k + (\cdots),$$

which we add as b_6 to the basis. We also add the pairs $(b_1, b_6), \ldots, (b_5, b_6)$ to P.

At this point, P contains 12 pairs, but it turns out that all corresponding S-polynomials reduce to zero, so we are done. The resulting Gröbner basis is $\{b_1, \ldots, b_6\}$. □

Algorithm 4.80 is not explicit about how to make the choice of $(p, q) \in P$ in line 4. Indeed, a different choice can lead to a different output. Every ideal I of $K[X_1, \ldots, X_n]$ has many different Gröbner bases (even for a fixed term order), and the only assertion about Algorithm 4.80 is that it will find one of them. We can eliminate this redundancy by imposing further constraints. A Gröbner basis G is called *reduced* if its elements are monic and we have $\operatorname{red}(g, G \setminus \{g\}) = g$ for all $g \in G$. It can be shown like in the commutative case that every ideal has exactly

one reduced Gröbner basis (for a prescribed term order). Starting from any finite Gröbner basis, e.g., some output of Algorithm 4.80, we can find the reduced Gröbner basis by first replacing every element g by $\mathrm{red}(g, G \setminus \{g\})$ and afterwards dividing every element from the left by its leading coefficient.

While the choices made in line 4 are irrelevant for the correctness and termination of the algorithm, they can have a strong influence on the runtime. Many people have thought about these choices, and have proposed several *selection strategies*. Common strategies select the next pair (p, q) based on $\max(\mathrm{lexp}(p), \mathrm{lexp}(q))$ (preferring lower ones), on the age (preferring older ones), or on the number of terms (preferring smaller ones). More sophisticated strategies compute a certain score for each pair in P and select the pair with the highest score. Another way to improve the performance is to identify *useless pairs*. There are certain criteria by which it is possible to detect at low cost whether a given pair $(p, q) \in P$ can be discarded without harming the correctness. Some of the criteria known from the commutative theory carry over to the noncommutative setting (Exercise 17), others don't (Exercise 16).

Yet another way to improve the performance is to handle several pairs at the same time. Instead of a single pair (p, q) in line 4, we can select a subset $S \subseteq P$ with $|S| \geq 1$ and set $P = P \setminus S$. In a preprocessing step, we then determine all the term-multiples of elements of B that may occur during the reduction of S-polynomials of the selected pairs. By a term-multiple we mean an operator of the form τb with τ being a term and $b \in B$. These term-multiples can be found by doing a "dry-run" of the reduction algorithm with coefficients replaced by Booleans that signal the potential nonzeroness of a coefficient. When a suitable collection of term-multiples has been constructed, we can determine all terms appearing in these term-multiples or the S-polynomials of the selected pairs. We then set up a matrix in which the columns are labeled by these terms, in decreasing order from left to right, and with two rows per selected pair and one row for each determined term-multiple. For each pair (p, q), we fill the row for p with the coefficients of $X^{\max(\mathrm{lexp}(p),\mathrm{lexp}(q))-\mathrm{lexp}(p)}\,\mathrm{lc}(p)^{-1}p$ and the row for q with the coefficients of $X^{\max(\mathrm{lexp}(p),\mathrm{lexp}(q))-\mathrm{lexp}(q)}\,\mathrm{lc}(q)^{-1}q$. For each term-multiple, we put its coefficients into the corresponding row. The resulting matrix is then brought into echelon form using Gaussian elimination, and we select all rows whose left-most nonzero entry is in a column that corresponds to a term which cannot be reduced by any element of B. There may be zero, one, or several such rows. For each of them, let h be the element of $K[X_1, \ldots, X_n]$ whose coefficient vector is the row and execute line 7 of Algorithm 4.80. Then continue with a new selection $S \subseteq P$ and repeat the procedure until $P = \emptyset$.

Example 4.83 Considering the same input as in the previous example, let us take as S the whole initial set $P = \{(b_1, b_2), (b_1, b_3), (b_2, b_3)\}$. The three S-polynomials corresponding to these pairs are differences of b_1, $S_k\,\mathrm{lc}(b_2)^{-1}b_2$, and $S_n\,\mathrm{lc}(b_3)^{-1}b_3$, and the terms occurring in these operators are $S_n^2S_k$, S_n^2, $S_nS_k^2$, S_nS_k, S_n, S_k^3, S_k^2, S_k, 1. For reducing polynomials involving these terms, we determine multiples of b_1, b_2, b_3 whose leading terms appear in this list.

They are $S_k b_3$ (with leading term $S_n S_k^2$), and b_1, b_2, b_3 themselves. These operators contain only terms that are already in the list, so we can form the matrix

$$\begin{array}{r} \\ b_1 \\ S_k \operatorname{lc}(b_2)^{-1} b_2 \\ S_n \operatorname{lc}(b_3)^{-1} b_3 \\ S_k b_3 \\ b_2 \\ b_3 \end{array} \begin{array}{c} \begin{array}{ccccccccc} S_n^2 S_k & S_n^2 & S_n S_k^2 & S_n S_k & S_n & S_k^3 & S_k^2 & S_k & 1 \end{array} \\ \begin{pmatrix} * & 0 & * & * & * & 0 & * & * & * \\ * & 0 & * & * & 0 & * & * & * & 0 \\ * & * & * & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & * & * & 0 & * & * & * & 0 \\ 0 & * & 0 & * & * & 0 & * & * & * \\ 0 & 0 & 0 & * & * & 0 & * & * & * \end{pmatrix} \end{array}.$$

Row reduction turns this matrix into

$$\begin{array}{r} \\ \\ \\ \\ \\ \rightarrow \\ \rightarrow \end{array} \begin{array}{c} \begin{array}{ccccccccc} S_n^2 S_k & S_n^2 & S_n S_k^2 & S_n S_k & S_n & S_k^3 & S_k^2 & S_k & 1 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & * & * & * \\ 0 & 1 & 0 & 0 & 0 & 0 & * & * & * \\ 0 & 0 & 1 & 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 1 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 1 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * & * \end{pmatrix} \end{array},$$

which has two rows (indicated by the arrows) that correspond to operators with new leading terms. The new leading terms are S_n and S_k^3, respectively, like for the operators b_4 and b_5 we found in the previous example. □

The theory of Gröbner bases can be generalized from ideals of $K[X_1, \ldots, X_n]$ to submodules of $K[X_1, \ldots, X_n]^d$. Elements of $K[X_1, \ldots, X_n]^d$ are K-linear combinations of module terms, a module term being an element of the form $X_1^{u_1} \cdots X_n^{u_n} e_i$ with $u_1, \ldots, u_n \in \mathbb{N}$ and e_i the ith unit vector. Term orders, reduction, the notion of Gröbner bases, and Buchberger's algorithm extend literally if we simply regard the unit vectors $e_1, \ldots, e_d$ as additional variables subject to the constraints $e_i e_j = 0$ for $i, j = 1, \ldots, d$. Of particular interest are term orders which first compare the index of the unit vector and then use a term order for the polynomials $X_1, \ldots, X_n$ for breaking ties. Such a term order is called a POT order (position over term), as opposed to a TOP order (term over position) which first looks at $X_1, \ldots, X_n$ and then uses the index of the unit vector to break ties.

One application of Gröbner bases for modules is the computation of syzygies and cofactors. A *syzygy* of $b_1, \ldots, b_m \in K[X_1, \ldots, X_m]$ is a vector $(p_1, \ldots, p_m) \in K[X_1, \ldots, X_n]^m$ with $p_1 b_1 + \cdots + p_m b_m = 0$. The set of all syzygies for a fixed choice of $b_1, \ldots, b_m$ forms a (left-)submodule of $K[X_1, \ldots, X_n]^m$, denoted by $\operatorname{Syz}(b_1, \ldots, b_m)$. Cofactors are the coefficients which are used in order to express an ideal element in terms of generators of an ideal: $f \in \langle b_1, \ldots, b_m \rangle$ means that there are $q_1, \ldots, q_m \in K[X_1, \ldots, X_n]$ with $f = q_1 b_1 + \cdots + q_m b_m$, and these $q_1, \ldots, q_m$ are called *cofactors* of f with respect to $b_1, \ldots, b_m$. They are in general not unique, but any two vectors of cofactors differ by a syzygy.

If $\{b_1, \ldots, b_m\}$ is a Gröbner basis, we can easily compute cofactors for any given $f \in \langle b_1, \ldots, b_m \rangle$ by an extended version of the reduction algorithm (Exercise 6). If it is not, we can compute a Gröbner basis $\{g_1, \ldots, g_k\}$ using Buchberger's algorithm. It is then easy to compute cofactors with respect to $g_1, \ldots, g_k$, but if we want cofactors with respect to the original basis, then we need to know how the elements of the Gröbner basis can be expressed in terms of the original basis elements. This information can be obtained by computing a Gröbner basis with respect to a POT order of the submodule generated by

$$\begin{pmatrix} b_1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} b_2 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots\ldots, \begin{pmatrix} b_m \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

The Gröbner basis will have the form

$$\begin{pmatrix} g_1 \\ q_{1,1} \\ q_{2,1} \\ \vdots \\ q_{m,1} \end{pmatrix}, \ldots, \begin{pmatrix} g_k \\ q_{1,k} \\ q_{2,k} \\ \vdots \\ q_{m,k} \end{pmatrix}, \begin{pmatrix} 0 \\ p_{1,1} \\ p_{2,1} \\ \vdots \\ p_{m,1} \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ p_{1,\ell} \\ p_{2,\ell} \\ \vdots \\ p_{m,\ell} \end{pmatrix},$$

where $\{g_1, \ldots, g_k\}$ is a Gröbner basis of $\langle b_1, \ldots, b_m \rangle$ and the matrix $((q_{i,j}))_{i=1,j=1}^{m,k} \in K[X_1, \ldots, X_n]$ is the basis change matrix that translates linear combinations of $g_1, \ldots, g_k$ into linear combinations of $b_1, \ldots, b_m$. Moreover, the vectors

$$(p_{1,j}, \ldots, p_{m,j}) \in K[X_1, \ldots, X_n]^m$$

form a basis of the syzygy module of $b_1, \ldots, b_m$.

With the help of the syzygy module, we can make the intersection of ideals constructive. Given two ideals $I = \langle b_1, \ldots, b_m \rangle$, $J = \langle d_1, \ldots, d_k \rangle$, an operator belongs to the intersection $I \cap J$ if and only if it can be written as a linear combination of the b_i and as a linear combination of the d_j. The search for operators $p_1, \ldots, p_m$ and $q_1, \ldots, q_k$ with $p_1 b_1 + \cdots + p_m b_m = q_1 d_1 + \cdots + q_k d_k$ is the same as the search for the syzygy module of $b_1, \ldots, b_m, -d_1, \ldots, -d_k$. Once we have a basis of the syzygy module, we can take the first m coordinates of each basis element and combine them with $b_1, \ldots, b_m$. The resulting operators are generators of the intersection ideal. This approach generalizes the idea of Exercise 14 in Sect. 4.2 to the case of several variables and provides an alternative way to compute an annihilating ideal for the sum of two D-finite objects from given annihilating ideals of the summands.

Gröbner bases for modules can also be used to uncouple systems of operator equations. Given a system $A \cdot f = 0$ with a known $A \in K[\partial]^{r\times r}$ and an unknown $f \in F^r$, we can consider the module generated in $K[\partial]^r$ by the rows of A. Computing a Gröbner basis of this module with respect to a POT order is equivalent to computing a Hermite normal form of A.

Exercises

1. Show that Hilbert's basis theorem, which says that every ideal of an Ore algebra $K[X_1, \ldots, X_n]$ has a finite basis, is equivalent to the ascending chain condition, which says that for every chain $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ of ideals of $K[X_1, \ldots, X_n]$ there exists an $m \in \mathbb{N}$ such that $I_m = I_{m+1} = \cdots$.

2. Let $\leq$ be a term order for the commutative polynomial ring $C[x_1, \ldots, x_n]$. Consider an Ore algebra $K[\partial_1, \ldots, \partial_n]$. Show that $\leq$ is also a valid term order for $K[\partial_1, \ldots, \partial_n]$ if we set $\partial_1^{u_1} \cdots \partial_n^{u_n} \leq \partial_1^{v_1} \cdots \partial_n^{v_n} \iff x_1^{u_1} \cdots x_n^{u_n} \leq x_1^{v_1} \cdots x_n^{v_n}$.

3. Let $\leq$ be a term order for the commutative ring $C[x_1, \ldots, x_n, y_1, \ldots, y_n]$.

a. Show that $\leq$ is also a valid term order for $C[x_1, \ldots, x_n, D_1, \ldots, D_n]$ (with the D_i being compared like the y_i).

b. Show that $\leq$ is also a valid term order for $C[x_1, \ldots, x_n, S_1, \ldots, S_n]$ (with the S_i being compared like the y_i).

4. Consider the Ore algebra $C[x, M_5]$ with the commutation rule $M_5 x = x^5 M_5$. Show that there is no term order for $C[x, M_5]$.

5. Show by an example that condition (iii) in the definition of term orders does not follow from conditions (i) and (ii) in general.

6. Extend Algorithm 4.76 so that it returns not only the remainder r but also cofactors $q_1, \ldots, q_m \in K[X_1, \ldots, X_n]$ such that $p - r = q_1 g_1 + \cdots + q_m g_m$ when applied to p and $G = \{g_1, \ldots, g_m\}$.

7⋆. Fill the gap in the proof of the implication $1 \Rightarrow 2$ of Theorem 4.77, i.e., show that if $G \subseteq K[X_1, \ldots, X_n]$ is such that every $u \in K[X_1, \ldots, X_n]$ has a unique remainder $\operatorname{red}(u, G)$, then we have $\operatorname{red}(u, G) = \operatorname{red}(u + c\tau \operatorname{lt}(g)^{-1} g, G)$ for every $c \in K$, every term τ, and every $g \in G$.

8. The set $G = \{D_x^3 - 5D_x^2 + 8D_x - 4, (4y + 1)D_x^2 - (16y + 4)D_x - yD_y + (15y + 4)\} \subseteq C(x, y)[D_x, D_y]$ is a Gröbner basis with respect to the lexicographic term order with $D_x < D_y$. Use Algorithm 4.73 to compute a Gröbner basis of $\langle G \rangle$ with respect to the lexicographic term order with $D_x > D_y$.

9⋆. Let $I \subseteq K[X_1, \ldots, X_n]$ be an ideal and G be a Gröbner basis of I. Show that the dimension of $K[X_1, \ldots, X_n]/I$ as a K-vector space is finite if and only if for every i there exists a $g \in G$ such that $\operatorname{lt}(g)$ is a power of X_i.

10. Let $I \subseteq K[X_1, \ldots, X_n]$ be an ideal and G be a Gröbner basis of I. Show that $1 \in I$ if and only if G contains an element of the form $u X_1^0 \cdots X_n^0$ with $u \in K \setminus \{0\}$.

11⋆. Let $I \subseteq K[X_1, \ldots, X_n]$, let $\preceq_1, \preceq_2$ be two term orders, and let G_1, G_2 be Gröbner bases of I with respect to $\preceq_1, \preceq_2$, respectively. Show that the number of terms τ with $\mathrm{red}_{\preceq_1}(\tau, G_1) = \tau$ is equal to the number of terms τ with $\mathrm{red}_{\preceq_2}(\tau, G_2) = \tau$. Must the sets of these terms also be equal?

12⋆. Show that $\{(1-k+n)S_n + (1+n), (1+k)S_k + 2(k-n)\} \subseteq C(n,k)[S_n, S_k]$ is a Gröbner basis and that $\{(1-k+n)S_n + (1+n), (1+k)S_k + 2(k+n)\}$ is not.

13. In line 2 of Algorithm 4.80, we do not initialize P with all pairs, as Theorem 4.79 suggests. Why is this okay?

14. A power series $a(x, y) \in C[[x, y]]$ has the annihilating ideal

$$\begin{aligned} I = \langle &(x-2)x D_x D_y + (2y+1)D_y^2 + (x+2)D_y, 4x D_x^2 D_y - (x+4y+4)D_x D_y^2 \\ &+ (y+1)D_y^3 - 2x D_x D_y + (4y+7)D_y^2 - 4x D_x + 4(y+3)D_y \rangle \\ &\subseteq C(x,y)[D_x, D_y]. \end{aligned}$$

Show that the series is D-finite.

15. A sequence $(a_{n,k})_{n,k=0}^{\infty}$ has the annihilating ideal

$$\begin{aligned} I = \langle &(1+k)(k+n)S_k + (k-n)(1+k+n), \\ &(1-k+n)(k+n)S_n - (1+n)(1+k+n) \rangle \subseteq C[n,k][S_n, S_k]. \end{aligned}$$

Show that the sequence is holonomic.

16. In the commutative case, we have

$$\min(\mathrm{lexp}(p), \mathrm{lexp}(q)) = 0 \Rightarrow \mathrm{red}(\mathrm{spol}(p,q), \{p,q\}) = 0.$$

Show that this does not work in the noncommutative case.

17⋆⋆. In the commutative case, we have the following criterion: whenever $B = \{b_1, \ldots, b_m\} \subseteq C[x_1, \ldots, x_n]$ and $p, u, v \in C[x_1, \ldots, x_n]$ are such that

1. there exist $q_1, \ldots, q_m \in C[x_1, \ldots, x_n]$ such that $\mathrm{spol}(u, p) = q_1 b_1 + \cdots + q_m b_m$ and $\mathrm{lt}(q_i b_i) \prec X^{\max(\mathrm{lexp}(u), \mathrm{lexp}(p))}$ for all i,
2. there exist $q_1, \ldots, q_m \in C[x_1, \ldots, x_n]$ such that $\mathrm{spol}(p, v) = q_1 b_1 + \cdots + q_m b_m$ and $\mathrm{lt}(q_i b_i) \prec X^{\max(\mathrm{lexp}(p), \mathrm{lexp}(v))}$ for all i, and
3. $\mathrm{lexp}(p) \leq \max(\mathrm{lexp}(u), \mathrm{lexp}(v))$,

then there also exist $q_1, \ldots, q_m \in C[x_1, \ldots, x_n]$ such that $\mathrm{spol}(u, v) = q_1 b_1 + \cdots + q_m b_m$ and $\mathrm{lt}(q_i b_i) \prec X^{\max(\mathrm{lexp}(u), \mathrm{lexp}(v))}$ for all i.

Show that this also works in the noncommutative case.

18. Let $G \subseteq K[X_1, \ldots, X_n]$ be a Gröbner basis and $g, h \in G$ with $g \neq h$. Show that $\operatorname{lexp}(g) \leq \operatorname{lexp}(h)$ implies that also $G \setminus \{h\}$ is a Gröbner basis of $\langle G \rangle$.

19. In the ideal

$$\begin{aligned} I = \langle &(x-1)D_x S_n + (x-1)D_x - (1+n)S_n + (1+n), \\ &(x-1)^2(x+1)D_x^2 + (x-1)(n+2x+nx)D_x - n(1+n)S_n + n(1+n)\rangle \\ &\subseteq C(n,x)[S_n, D_x], \end{aligned}$$

find an element which is a $C(n, x)$-linear combination of powers of $S_n D_x$.

20. In the commutative case, an alternative way to compute the intersection of two ideals $I = \langle b_1, \ldots, b_m \rangle$ and $J = \langle d_1, \ldots, d_k \rangle$ of $C[x_1, \ldots, x_n]$ is to compute the elimination ideal

$$\langle tb_1, \ldots, tb_m, (1-t)d_1, \ldots, (1-t)d_k \rangle \cap C[x_1, \ldots, x_n],$$

where t is an additional variable. Does this also work in the noncommutative case?

References

Gröbner bases for ideals of a commutative polynomial ring $C[x_1, \ldots, x_n]$ were introduced by Buchberger in his Ph.D. thesis [117]. They play a central role in computer algebra. Introductory texts on the subject include [29, 47, 167].

A starting point for the development of Gröbner bases for non-commutative rings was the influential paper of Bergman from 1978 [56], who considered the case of a free algebra. In general, an ideal in such a ring need not have a finite Gröbner basis, so that during the 1980s, various people have investigated theories for Gröbner bases in more special non-commutative settings, including Galligo [200], Mora [332], Apel and Lassner [37], Takayama [422], Kandri-Rody and Weispfenning [256]. An introductory article of Mora [333] contains a proof of the non-commutative chain criterion (Exercise 17).

Zeilberger did not use Gröbner basis in his landmark paper [468], but suggested the use of Gröbner bases in this context. Chyzak and Salvy took up this suggestion and introduced Gröbner bases for Ore algebras [153, 154, 157].

The idea to view the reduction process from the perspective of linear algebra goes back to Lazard [305] and culminated in Faugère's F4 algorithm [189]. These techniques were developed for the commutative case but the extension to the non-commutative setting is straightforward. Less straightforward is the extension of another idea for speeding up Gröbner basis computation, which was introduced by Faugère under the name F5 [190] and has led to the development of so-called signature-based Gröbner bases algorithms [182]. A signature-based Gröbner bases algorithm for the non-commutative case was worked out by Sun, Wang, Ma, and Zhang [420].