5 Keeping the Data Small: Modular Methods

5.1 Modular gcd of Polynomials in $\mathbb{Z}[x]$

First of all we note the following important fact:

Lemma 5.1 (Gauss) *For any $f, g \in \mathbb{Z}[x]$ (not both zero) we have*

$$\begin{aligned} \operatorname{cont}(fg) &= \operatorname{cont}(f)\ \operatorname{cont}(g), \\ \operatorname{pp}(fg) &= \operatorname{pp}(f)\ \operatorname{pp}(g). \end{aligned}$$

Proof It suffices to show that the product of two primitive polyniomials is itself primitive. Let $u, v \in \mathbb{Z}[x]$ be primitive, i.e., $\operatorname{cont}(u) = \operatorname{cont}(v) = 1$. If $\operatorname{cont}(uv) \neq 1$, i.e., uv is not primitive, then there is a prime p that divides all the coefficients of uv. Set

$$\begin{aligned} u &= u_m x^m + u_{m-1} x^{m-1} + \cdots + u_0, \\ v &= v_n x^n + v_{n-1} x^{n-1} + \cdots + v_0. \end{aligned}$$

Since u, v are primitive there are indices j, k such that $p \not| u_j$ and $p \not| v_k$. Note that the coefficient of x^{j+k} in uv is

$$u_j v_k + u_{j+1} v_{k-1} + \cdots + u_{j+k} v_0 + u_{j-1} v_{k+1} + \cdots + u_0 v_{k+j}$$

Choose j, k to be as small as possible; then the fist summand above is not divisible by p while all the rest are. It follows that the coefficient as a whole is not divisible by p which is a contradiction. □

In fact the lemma holds if $\mathbb{Z}$ is replaced with any unique factorization domain; indeed the same proof goes through. An immediate consequence of the lemma is the following.

Lemma 5.2 *For any $f, g \in \mathbb{Z}[x]$ (not both zero) we have*

$$\begin{aligned} \operatorname{cont}(\gcd(f, g)) &= \gcd(\operatorname{cont}(f), \operatorname{cont}(g)), \\ \operatorname{pp}(\gcd(f, g)) &= \gcd(\operatorname{pp}(f), \operatorname{pp}(g)). \end{aligned}$$

Proof Let $h = \gcd(f, g)$ so that we have $f = \hat{f}h$ and $g = \hat{g}h$. Of course $\gcd(\hat{f}, \hat{g}) = 1$ since $\gcd(\hat{f}, \hat{g})h$ divides both f and g and so it divides their gcd, i.e., it divides h. It follows that $\gcd(\operatorname{cont}(\hat{f}), \operatorname{cont}(\hat{g})) = 1$ and $\gcd(\operatorname{pp}(\hat{f}), \operatorname{pp}(\hat{g})) = 1$ (because both of these divide $\gcd(\hat{f}, \hat{g})$).

By Gauss' Lemma, $\operatorname{cont}(f) = \operatorname{cont}(\hat{f})\operatorname{cont}(h)$ and $\operatorname{cont}(g) = \operatorname{cont}(\hat{g})\operatorname{cont}(h)$. It follows that $\gcd(\operatorname{cont}(f), \operatorname{cont}(g)) = \operatorname{cont}(h)$ since $\gcd(\operatorname{cont}(\hat{f}), \operatorname{cont}(\hat{g})) = 1$.

Similarly, $\operatorname{pp}(f) = \operatorname{pp}(\hat{f})\operatorname{pp}(h)$ and $\operatorname{pp}(g) = \operatorname{pp}(\hat{g})\operatorname{pp}(h)$. Since $\gcd(\operatorname{pp}(\hat{f}), \operatorname{pp}(\hat{g})) = 1$ it follows that $\gcd(\operatorname{pp}(f), \operatorname{pp}(g)) = \operatorname{pp}(h)$. □

It follows that we may restrict our attention to primitive polynomials and the result will always be a primitive polynomial. Another useful observation is that

$$\operatorname{lc}(\gcd(f, g)) | \gcd(\operatorname{lc}(f), \operatorname{lc}(g))$$

(prove this). If $f = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_0$ and p is a positive integer then we define

$$(f \bmod p) = (a_m \bmod p)x^m + (a_{m-1} \bmod p)x^{m-1} + \cdots + (a_0 \bmod p)$$

which is a polynomial in $\mathbb{Z}_p[x]$. During this section, we abbreviate $(f \bmod p)$ to f_p. Note that if $p \not| \operatorname{lc}(f)$ then $\deg(f_p) = \deg(f)$; the converse is also true.

Exercise 5.1 *Redo the proof of Lemma 5.1 but using $u \bmod p$ and $v \bmod p$ to derive a contradiction.*

Let us take another look at the problem of finding the gcd of

$$A = x^8 + x^6 - 3x^4 - 3x^3 + 8x^2 + 2x - 5,$$
$$B = 3x^6 + 5x^4 - 4x^2 - 9x + 21.$$

Put

$$A = PH, \quad B = QH,$$

where $H = \gcd(A, B)$. Consider these equations modulo 5, i.e., view them as holding in the ring $\mathbb{Z}_5[x]$. By computing in $\mathbb{Z}_5[x]$ we see that $\gcd(A_5, B_5) = 1$. We have $5 \not| \operatorname{lc}(H)$ since $5 \not| \gcd(\operatorname{lc}(A), \operatorname{lc}(B))$ and so $\deg(H_5) = \deg(H)$. Thus $\deg(H) \leq \deg(\gcd(A_5, B_5)) = 0$ so that $\deg(H) = 0$. Thus $H = 1$ and so $\gcd(A, B) = 1$.

Suppose we want the gcd of two arbitrary primitive polynomials of $\mathbb{Z}[x]$. Let $p_1, \ldots, p_k$ be a sequence of primes. For each i we can compute the gcd modulo p_i, i.e., compute in $\mathbb{Z}_{p_i}[x]$. The goal is to combine these results to obtain the result in $\mathbb{Z}[x]$. (As an extreme case we could compute with only one prime which is so large that $\gcd(A_p, B_p) = \gcd(A, B)$ but this is of little advantage.) Note that by computing in $\mathbb{Z}_p[x]$ for some prime p of reasonable size we avoid completely the problem of intermediate expression swell as far as coefficients are concerned. Clearly it is well worth trying to develop a method along these lines. We have three problems to address:

1. How do we combine the various results in the $\mathbb{Z}_{p_i}[x]$ into a single result in $\mathbb{Z}[x]$?

2. Given $A, B \in \mathbb{Z}[x]$ how big can the coefficients of $\gcd(A, B)$ be? We need to know this because we intend to work with modular arithmetic and then recover integer coefficients. To put it simply suppose we are told that an integer a taken modulo 17 gives 15. All we can deduce from this information is that $a = 17q + 15$ for some integer q. However if we are also told that $|a| < 8$ then $-8 < a < 8$ and we can deduce that $a = -2$. We discuss this point in more detail below.

3. Which primes should we choose? Are there any that should be avoided?

We illustrate the preceding points with a more detailed example. Take

$$A = 3x^4 + 4x^3 - 6x^2 - 3x + 2,$$
$$B = 9x^5 + 21x^4 + 6x^3 + x^2 + x - 2.$$

Put

$$H = \gcd(A, B).$$

Note that A, B are primitive so that H must be primitive. Also $\deg(H) \leq \min(\deg(A), \deg(B)) = 4$. Moreover an easy computation shows that $A \not| B$ and so $\deg(H) < 4$, thus we may put

$$H = h_3x^3 + h_2x^2 + h_1x + h_0.$$

Our aim is to work modulo p where p is a prime (possibly using several p) and compute $\gcd(A_p, B_p)$ using Euclid's algorithm in $\mathbb{Z}_p[x]$ (since p is prime it follows that $\mathbb{Z}_p$ is a field and Euclid's algorithm applies). If we are lucky then the gcd thus obtained is equal to H_p (actually we have to modify this a little, as we shall see near the end of the example). We are not always lucky, e.g., x and $x + 5$ are coprime so their gcd is 1 but taken modulo 5 their gcd is x. Notice that if $p \not| \operatorname{lc}(A)$ or $p \not| \operatorname{lc}(B)$ then $p \not| \operatorname{lc}(H)$ and so for such a p we have $\deg(\gcd(A_p, B_p)) \geq \deg(H)$. Moreover if $\deg(\gcd(A_p, B_p)) > 3$ then we know that $\gcd(A_p, B_p) \neq H_p$ and so we must reject p as 'unlucky'.

Now working modulo 2 we have

$$A_2 = x^4 + x,$$
$$B_2 = x^5 + x^4 + x^2 + x,$$

and from Euclid's algorithm (in the ring $\mathbb{Z}_2[x]$) we have

$$\gcd(A_2, B_2) = x^4 + x.$$

This shows that there must be something wrong with 2 as a modulus. The next prime, 3, is also a bad choice because it divides $\operatorname{lc}(A)$ and $\operatorname{lc}(B)$. Now working modulo 5 we have

$$A_5 = 3x^4 + 4x^3 + 4x^2 + 2x + 2,$$
$$B_5 = 4x^5 + x^4 + x^3 + x^2 + x + 3,$$

and

$$F_5 = \gcd(A_5, B_5) = x^3 + 4x^2 + 2x + 1.$$

This shows that 5 might be acceptable in the sense that $F_5 = H_5$ so we proceed under this assumption. First of all we view F_5 as an element of $\mathbb{Z}[x]$ and check to see if $F_5|A$ and $F_5|B$ for if this is so then $H = F_5$ and we have found the answer. Unfortunately $F_5 \not| A$. This does not mean that 5 was a bad choice: it *might* be a bad choice *or* (if we are lucky) we have not yet recovered the coefficients of H completely because at least one of them has been 'collapsed' by taking it modulo 5. We now do the same with the next prime to obtain

$$F_7 = \gcd(A_7, B_7) = x^3 + 5x + 4,$$

and $F_7 \not| A$. Assuming that both 5 and 7 are good choices of moduli we now have the following four pairs of simultaneous congruences:

$$\begin{array}{llll} h_3 \equiv 1 \pmod 5, & & h_3 \equiv 1 \pmod 7, \\ h_2 \equiv 4 \pmod 5, & & h_2 \equiv 0 \pmod 7, \\ h_1 \equiv 2 \pmod 5, & & h_1 \equiv 5 \pmod 7, \\ h_0 \equiv 1 \pmod 5, & & h_0 \equiv 4 \pmod 7. \end{array}$$

Let us find the possible solutions to the last pair of congruences. The first congruence shows that $h_0 = 1 + 5q$ for $q \in \mathbb{Z}$. Substituting this into the second congruence we obtain:

$$5q \equiv 3 \pmod 7.$$

Now

$$3 \cdot 5 - 2 \cdot 7 = 1.$$

(The numbers 3 and -2 can be obtained from the Extended Euclidean Algorithm applied to 5 and 7.) Thus

$$3 \cdot 5 \equiv 1 \pmod 7$$

i.e., 3 is the multiplicative inverse of 5 in the field $\mathbb{Z}_7$ (remember that a congruence modulo p is the same as an equation in the field $\mathbb{Z}_p$). We may now deduce that

$$\begin{aligned} q &\equiv 3 \cdot 3 \pmod 7, \\ &\equiv 2 \pmod 7. \end{aligned}$$

So for a simultaneous solution we take $q = 2 + 7q'$ in $1 + 5q$ which gives us $11 + 35q'$. What we have now is that $h_0 \equiv 11 \pmod{35}$. Doing the same for the other pairs of congruences we obtain

$$F_{35} = x^3 + 14x^2 + 12x + 11$$

as the candidate for H_{35}. (Notice that we now have a candidate with modulus 35 even though we have only carried out gcd computations with the moduli 5 and 7.) If all the coefficients of H are in the range $-17 < h \leq 18$ then we already have H and not just H_{35}. A simple calculation shows that $F_{35} \not| A$. But before giving up we should re-examine one crucial point in the preceding calculations. When finding $\gcd(A_5, B_5)$ and $\gcd(A_7, B_7)$ we produced *monic* polynomials for the results. These are perfectly valid but then so is any non-zero constant multiple of them! When calculating $\gcd(A_5, B_5)$ we are really trying to obtain H_5 and the leading coefficient of H might not be 1 so that the leading coefficient of H_5 need not be 1 either. If we knew $\mathrm{lc}(H)$ there would be no problem: we would simply find the monic gcd as before and then multiply it with $\mathrm{lc}(H) \bmod 5$. (Note here an important consequence of the fact that we use only primes p that do not divide $\gcd(\mathrm{lc}(A), \mathrm{lc}(B))$. For such primes we have that $\mathrm{lc}(H) \bmod p$ is not 0.) Unfortunately we do not know $\mathrm{lc}(H)$ but we *do* know that $\mathrm{lc}(H) | c$ where $c = \gcd(\mathrm{lc}(A), \mathrm{lc}(B)) = 3$ (equality need not hold—consider $(x+1)(2x+3)$ and $(x+1)(2x+5)$). It follows that if we multiply F_5 with $(c \bmod 5)$ *in the ring* $\mathbb{Z}_p[x]$ then the result, interpreted as a polynomial from $\mathbb{Z}_p[x]$, is dH_5 for some non-zero constant d. Similarly for F_7. This replaces F_5, F_7 with

$$\begin{aligned} F_5^* &= 3x^3 + 2x^2 + x + 3, \\ F_7^* &= 3x^3 + x + 5. \end{aligned}$$

These now yield the candidate

$$F_{35}^* = 3x^3 + 7x^2 + x - 2.$$

If we are lucky then when this is viewed as an element of $\mathbb{Z}[x]$ it is just eH for some constant e. Since H must be primitive all we have to do in order to eliminate e is to take the primitive part of F_{35}^*. Clearly F_{35}^* is primitive and in fact it divides both A and B so that it is the required gcd.

Exercise 5.2 *We obtained F_{35}^* from F_5^* and F_7^*. Note that, in fact, we obtain the same result from $3F_{35}$ (after reducing all coefficients modulo 35). Is this just a coincidence?*

5.2 The Chinese Remainder Problem

Suppose that D is an integral domain in which a version of the Euclidean algorithm holds (such rings are called *Euclidean domains*—it can be shown that every Euclidean domain is a UFD). For our applications D is either $\mathbb{Z}$ or $k[x]$ where k is a field. We are given remainders $r_1, \ldots, r_n \in D$ and moduli $m_1, \ldots, m_n \in D - \{0\}$ which are pairwise coprime. (Two elements are said to be *coprime* if their gcd is 1. For integers this means literally that the gcd is the number 1 while for polynomials this means that the gcd is a constant.) The problem is to find $r \in D$ such that

$$r \equiv r_i \pmod{m_i}$$

for $1 \leq i \leq n$. For simplicity let $n = 2$. We have

$$r \equiv r_1 \pmod{m_1} \tag{1}$$
$$r \equiv r_2 \pmod{m_2} \tag{2}$$

Every solution of (1) has form $r_1 + \sigma m_1$. So we have to find σ such that $r_1 + \sigma m_1 \equiv r_2 \pmod{m_2}$. We have $\gcd(m_1, m_2) = 1 = cm_1 + dm_2$ and so $cm_1 \equiv 1 \pmod{m_2}$, where c can be computed by the Extended Euclidean algorithm. (In more highbrow notation we write $c = m_1^{-1}$ in the ring of remainders modulo m_2.) Now choose $\sigma = c(r_2 - r_1) \pmod{m_2}$. Thus

$$\begin{aligned} r_1 + \sigma m_1 &\equiv r_1 + c(r_2 - r_1)m_1 \pmod{m_2} \\ &\equiv r_1 + r_2 - r_1 \pmod{m_2}. \end{aligned}$$

Note that the preceding argument applies to any Euclidean domain.

Theorem 5.1 *The Chinese Remainder Problem always has a solution which can be computed by the algorithm CRA_2 given below.*

Algorithm: $CRA_2(r_1, r_2, m_1, m_2) \mapsto r$

1. $c := m_1^{-1} \bmod m_2$;

2. $r_1' := r_1 \bmod m_1$;

3. $\sigma := c(r_2 - r_1') \bmod m_2$;

4. $r := r_1' + \sigma m_1$;

We note that the output r of the algorithm has the property that the simultaneous congruences

$$x \equiv r_1 \pmod{m_1}$$
$$x \equiv r_2 \pmod{m_2}$$

hold for x if and only if

$$x \equiv r \pmod{m_1 m_2}.$$

For if x satisfies the two congruences then we have

$$x \equiv r \pmod{m_1}$$
$$x \equiv r \pmod{m_2}$$

so that $m_1 \mid x - r$ and $m_2 \mid x - r$. Now the fact that $\gcd(m_1, m_2) = 1$ implies that $m_1 m_2 \mid x - r$ so that $x \equiv r \pmod{m_1 m_2}$. On the other hand if $x \equiv r \pmod{m_1 m_2}$ then $x \equiv r \pmod{m_i}$ for $i = 1, 2$ and the result follows since $r \equiv r_i \pmod{m_i}$ for $i = 1, 2$.

We can solve CRP_n (i.e., the Chinese Remainder Problem with n remainders) by applying CRA_2 recursively: given the problem

$$\begin{aligned} r &\equiv r_1 \pmod{m_1} \\ r &\equiv r_2 \pmod{m_2} \\ r &\equiv r_3 \pmod{m_3} \\ &\vdots \end{aligned}$$

we solve the first two congruences and obtain r_{12} as the answer. The problem now reduces to solving

$$\begin{aligned} r &\equiv r_{12} \pmod{m_1 m_2} \\ r &\equiv r_3 \pmod{m_3} \\ &\vdots \end{aligned}$$

Just as for the case of two remainders we have that

$$x \equiv r_i \pmod{m_i}, \quad 1 \leq i \leq n,$$

if and only if

$$x \equiv r \pmod{m_1 m_2 \cdots m_n}.$$

This explains the utility of the Chinese Remainder Theorem. We can work with conveniently sized moduli $m_1, \ldots, m_n$ and then construct the result we would get by working with the single large modulus $m_1 m_2 \cdots m_n$.

It is now fairly easy to deduce:

Theorem 5.2 *For the case $D = \mathbb{Z}$ the solution r computed by CRA_n is bounded as follows*

$$0 \leq r < m_1 m_2 \cdots m_n.$$

Moreover there is exactly one such r.

Theorem 5.3 *For the case $D = k[x]$ the solution $r(x)$ computed by CRA_n is either 0 or bounded in degree as follows*

$$\deg(r) < \deg(m_1) + \cdots + \deg(m_n).$$

Moreover there is exactly one such $r(x)$.

To sum up, stated purely as a theorem we have:

Theorem 5.4 (Chinese Remainder Theorem for the Integers) *Assume $r_1, r_2 \ldots, r_n \in \mathbb{Z}$ and $m_1, m_2, \ldots, m_n \in \mathbb{Z}$ where $m_i > 1$, for $1 \leq i \leq n$, and m_i, m_j are comprime (i.e., $\gcd(m_i, m_j) = 1$) for $1 \leq i < j \leq n$. Then there is an integer x such that*

$$\begin{aligned} x &\equiv r_1 \pmod{m_1} \\ x &\equiv r_2 \pmod{m_2} \\ &\vdots \\ x &\equiv r_n \pmod{m_n}. \end{aligned}$$

Moreover setting $M = m_1 m_2 \cdots m_n$ we have that $x + qM$ is also a solution for all $q \in \mathbb{Z}$ and all solutions are of this form.

Exercise 5.3 *Suppose that we have two moduli which are not coprime but still wish to solve the two simultaneous congruences as in the case when m_1, m_2 are coprime. Find necessary and sufficient conditions for this to be possible and give an algorithm.*

In fact we can give a direct solution to the general case of the problem as follows. Let $M_i = m_1 m_2 \cdots m_{i-1} m_{i+1} \cdots m_n$ for $1 \leq i \leq n$. Find $b_1, b_2, \ldots, b_n$ such that

$$b_i M_i \equiv 1 \pmod{m_i},$$

for $1 \leq i \leq n$ (the b_i exist because $\gcd(M_i, m_i) = 1$). Then x is a solution to the system

$$\begin{aligned} x &\equiv r_1 \pmod{m_1} \\ x &\equiv r_2 \pmod{m_2} \\ &\vdots \\ x &\equiv r_n \pmod{m_n} \end{aligned}$$

if and only if

$$x \equiv r_1 b_1 M_1 + r_2 b_2 M_2 + \cdots + r_n b_n M_n \pmod{M}.$$

We now take a closer look at the integer case. In applications we want to find one or more integers by calculating with several moduli and then combining the results. How do we choose the moduli? Of course they must be coprime—in many applications we use primes for the moduli so the coprimeness condition follows automatically. The solution r which we recover from the residues satisfies

$$0 \leq r < M,$$

where $M = m_1 m_2 \cdots m_n$. Because we are trying to recover an integer which might be positive or negative it is much more convenient to write this inequality as

$$-M/2 < r' \leq M/2,$$

where

$$r' = \begin{cases} r, & \text{if } r \leq M/2; \\ r - M & \text{if } r > M/2. \end{cases}$$

Note that r' is also a solution to the Chinese Remainder Problem since $r \equiv r - M \pmod{M}$. Suppose that the integer we are trying to recover is R. Just as in the discussion in §5.1, all we can deduce so far is that $R = qM + r$ for some integer q. However if we also know that $|R| < M/2$ then we may immediately deduce that $R = r'$. Thus if we have an upper bound B for $|R|$ then we simply need to ensure that the moduli are chosen so that $M > 2B$. To put it another way if $M > 2B$ then in the range $[-B, B]$ there is exactly one symmetric remainder modulo M since if $-B \leq r \leq B$ then $r - M < r - 2B \leq B - 2B = -B$ while $r + M > r + 2B \geq -B + 2B = B$. Typically B is fairly large (after all the reason for going to all this trouble is because R is large and we want to avoid arithmetic with large integers). We therefore have to strike a balance between small moduli, in which case we will need very many of them, and moduli which are so large that we gain little

if any advantage in the arithmetic. Typically the moduli are chosen as large as possible provided they fit into a word of memory.

The preceding discussion also leads us to forego the preference induced by most mathematics texts (and, so far, these notes): in discussing arithmetic modulo m the remainders r are chosen in the range $0 \leq r < m$. For our applications of the Chinese Remainder Theorem it is more natural to choose the remainders in the range $-m/2 < r \leq m/2$. This latter representation is called *symmetric*. (Maple allows the user to switch to the symmetric representation by the assignment `‘mod‘:=mods`, the standard representation, which is the default, can be regained by `‘mod‘:=modp`.)

For an interesting account of the various guises of the Chinese Remainder Theorem see P. J. Davies and R. Hersh [22] (the book as a whole is worth reading). Here we simply note that in one form it expresses an isomorphism of rings which, in the integer case, is:

$$\mathbb{Z}_{m_1 m_2} \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2},$$

when m_1, m_2 are coprime (for the meaning of the right hand side see Exercise 4.17). It is only fair to point out that the Greek mathematician Nikomachos gave a version of the theorem at about the same time as Sun-Tsu who is normally credited with it (both date from the first century A.D.)

Finally, the Chinese Remainder Theorem and associated algorithm are fundamental to various applications in computer algebra. It is therefore essential that you understand fully both the result and its applications.

5.3 Bound on the Coefficients of the gcd

Let $A, B \in \mathbb{Z}[x]$. It is tempting to conjecture that the coefficients of $\gcd(A, B)$ can be no larger in absolute value than the largest absolute value of the coefficients of A or B. Unfortunately this is not the case, for example consider

$$\begin{array}{rcccl} A & = & x^3 + x^2 - x - 1 & = & (x+1)^2(x-1); \\ B & = & x^4 + x^3 + x + 1 & = & (x+1)^2(x^2 - x + 1); \\ \gcd(A,B) & = & x^2 + 2x + 1 & = & (x+1)^2. \end{array}$$

Clearly the problem of finding bounds for $\gcd(A, B)$ is solved if we can find bounds for the coefficients of the divisors of a given polynomial; we proceed to address this question for polynomials from $\mathbb{C}[x]$ (we will give an overview of the situation, details can be found in [67]).

Let

$$P = p_0 x^d + p_1 x^{d-1} + \cdots + p_d,$$

where $p_0 \neq 0$ (note the indexing of the coefficients; this makes subsequent expressions a little simpler). Let $\alpha_1, \alpha_2, \ldots, \alpha_d$ be all the d complex roots of P so that

$$P = p_0(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_d), \tag{6}$$

some roots might be repeated of course. Recall that for a complex number $\alpha = a + ib$, its absolute value is defined by $|\alpha| = \left(a^2 + b^2\right)^{1/2}$. We define two measures as follows.

$$||P|| = \left(|p_0|^2 + |p_1|^2 + \cdots + |p_d|^2\right)^{1/2},$$

$$M(P) = |p_0| \prod_{i=1}^{d} \max(1, |\alpha_i|).$$

The second measure has the very useful, and obvious, property that it is multiplicative, i.e., $M(P_1P_2) = M(P_1)M(P_2)$. It follows that if P and Q are monic (i.e., their leading coefficients are 1) and P divides Q then $M(P) \leq M(Q)$. Although $M(P)$ is difficult to compute it can be expressed as

$$M(P) = \exp\left(\int_0^1 \log|P(e^{2\pi it})|\,dt\right). \tag{7}$$

From this it is then possible to show that

$$M(P) \leq ||P||, \tag{8}$$

which was first proved by Landau (1905) by different methods. Now considering the expansion of the r.h.s. of (6) it is easy to show that

$$|p_i| \leq \binom{d}{i} M(P),$$

for $0 \leq i \leq d$. The last two inequalities can now be used to derive the desired bound on the coefficients of a divisor of P.

Theorem 5.5 *Let $Q = q_0x^r + q_1x^{r-1} + \cdots + q_r$ be a polynomial in $\mathbb{C}[x]$ that divides P. Then*

$$|q_i| \leq \binom{r}{i} \frac{|q_0|}{|p_0|} ||P||$$

for $0 \leq i \leq r$.

Proof We have

$$|q_i| \leq \binom{r}{i} M(Q).$$

If P and Q are monic, we have $M(Q) \leq M(P)$ and the claim follows from (8).

Finally if P or Q is not monic we proceed by observing that $p_0^{-1}P$ is monic and divides the monic polynomial $q_0^{-1}Q$. □

This result was observed by Mignotte [44]. Note that we can simplify the r.h.s. of the inequality to obtain

$$|q_i| \leq 2^r \frac{|q_0|}{|p_0|} ||P|| \leq 2^d \frac{|q_0|}{|p_0|} ||P||,$$

(justify the first inequality; there is a very simple reason). Finally if $P, Q \in \mathbb{Z}[x]$ then of course $|q_0| \leq |p_0|$ so that we finally obtain the inequality

$$|q_i| \leq 2^d ||P||.$$

We will refer to this as the *Landau-Mignotte inequality.*

Exercise 5.4 *Prove that if $f(x)$ is non-constant and (Riemann) integrable then*

$$\exp\left(\int_0^1 \log f(x)\,dx\right) \leq \int_0^1 f(x)\,dx. \tag{9}$$

For this use the fact that if $a_1, a_2, \ldots, a_n$ are numbers then

$$\frac{a_1 + a_2 + \cdots a_n}{n} \leq (a_1 a_2 \cdots a_n)^{1/n}.$$

Remember that an integral is the limit of a sum. Use (9) to derive (8) from (7). (Warning: this exercise is quite hard.)

Exercise 5.5 *Let $A = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_0$ and $B = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_0$ be polynomials in $\mathbb{Z}[x]$, where $m = \deg(A)$ and $n = \deg(B)$. Show that the absolute value of each coefficient of $\gcd(A, B)$ is bounded by*

$$2^{\min(m,n)} \gcd(a_m, b_n) \min\left(\frac{1}{|a_m|} \left(\sum_{i=0}^{m} a_i^2 \right)^{1/2}, \frac{1}{|b_n|} \left(\sum_{i=0}^{n} b_i^2 \right)^{1/2} \right).$$

5.4 Choosing Good Primes

There is a problem with the modular approach: given $A, B \in \mathbb{Z}[x]$ we clearly must choose a prime p that does not divide both $\mathrm{lc}(A)$ and $\mathrm{lc}(B)$. Now suppose that

$$A = PG, \quad B = QG.$$

Recall that for a polynomial F, F_p denotes $(F \bmod p)$. We have

$$A_p = P_p G_p, \quad B_p = Q_p G_p.$$

Unfortunately G_p might not be the gcd of A_p, B_p modulo p. For example take

$$A = x - 3, \quad B = x + 2, \quad p = 5.$$

Clearly $\gcd(A, B) = 1$ but

$$A_5 = B_5 = x + 2$$

so that $\gcd(A_5, B_5) = x + 2$.

Lemma 5.3 *Let $A, B \in \mathbb{Z}[x]$ and p a prime that does not divide both $\mathrm{lc}(A)$, $\mathrm{lc}(B)$. Then*

$$\deg(\gcd(A_p, B_p)) \geq \deg(\gcd(A, B)).$$

Proof $\gcd(A, B)_p$ divides both A_p and B_p and so it divides $\gcd(A_p, B_p)$. Now $\deg(\gcd(A_p, B_p)) \geq \deg(\gcd(A, B)_p)$ but $p \not| \mathrm{lc}(\gcd(A, B))$ and so $\deg(\gcd(A, B)_p) = \deg(\gcd(A, B))$. □

Let us call a prime p which doesn't work *unlucky*, i.e., $\gcd(A_p, B_p) \neq \gcd(A, B)_p$ (to be strictly accurate what we mean here is that $\gcd(A_p, B_p) \neq c \gcd(A, B)_p$ for any constant c—we drop c under the convention that in $\mathbb{Z}_p[x]$ we normalize polynomials to have 1 as leading coefficient when we are thinking of gcd's). Suppose that we have two primes p_1, p_2 which meet the assumption of the last lemma. If $\deg(\gcd(A_{p_1}, B_{p_1})) > \deg(\gcd(A_{p_2}, B_{p_2}))$ then we can immediately deduce that p_1 is unlucky (why?). How many unlucky primes are there? In order to answer this question

we need a faithful old workhorse of nineteenth century elimination theory which was introduced by J. Sylvester. Put

$$A = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_0,$$
$$B = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_0.$$

We assume that neither of these polynomials is 0. The *resultant* of A, B is

$$\mathrm{Res}(A,B) = \begin{vmatrix}
a_m & a_{m-1} & \dots & a_0 & & & & & \\
 & a_m & a_{m-1} & \dots & a_0 & & & & \\
 & & \cdot & & & \cdot & & & \\
 & & & \cdot & & & \cdot & & \\
 & & & & \cdot & & & \cdot & \\
 & & & & & \cdot & & & \cdot \\
 & & & & & & a_m & a_{m-1} & \dots & a_0 \\
b_n & b_{n-1} & \dots & b_0 & & & & & \\
 & b_n & b_{n-1} & \dots & b_0 & & & & \\
 & & \cdot & & & \cdot & & & \\
 & & & \cdot & & & \cdot & & \\
 & & & & \cdot & & & \cdot & \\
 & & & & & \cdot & & & \cdot \\
 & & & & & & b_n & b_{n-1} & \dots & b_0
\end{vmatrix}$$

where there are n rows of a-entries, m rows of b-entries and the blank spaces consist of 0. In the preceding we allow the possibility that one of a_m, b_n might be 0. Strictly speaking we should thus use a notation such as $\mathrm{Res}_{m,n}(A.B)$ but m, n are always understood from the context.

Theorem 5.6 *Suppose that $a_m \neq 0$ or $b_n \neq 0$. Then A and B have a non-constant common factor if and only if* $\mathrm{Res}(A,B) = 0$. *(Recall that we assumed $A \neq 0$ and $B \neq 0$.)*

Proof First of all we

CLAIM: A and B have a non-constant common factor if and only if there are polynomials ϕ and ψ, at least one of which is non-zero and of degrees strictly less than m and n respectively, such that $\psi A = \phi B$.

The proof is simple: if A and B do have a non-constant common factor H then we may put $A = \phi H$, $B = \psi H$ so that $\psi A = \phi B$ and clearly $\deg(\phi) < \deg(A)$, $\deg(\psi) < \deg(B)$. Conversely, suppose that $\psi A = \phi B$ and consider the irreducible factors of B. By the uniqueness of factorization each irreducible factor of B must appear amongst the irreducible factors of ψA (up to a constant multiple). The corresponding factors from ψA cannot all come from ψ because $\deg(\psi) < \deg(B)$. This completes the proof of the Claim.

Now put

$$\phi = \alpha_m x^{m-1} + \cdots + \alpha_1,$$
$$\psi = \beta_n x^{n-1} + \cdots + \beta_1.$$

When can we have $\psi A = \phi B$? Multiplying the two sides out and equating coefficients of corresponding powers of x we see that this is equivalent to:

$$\begin{aligned} a_0\beta_1 &= b_0\alpha_1, \\ a_1\beta_1 + a_0\beta_2 &= b_1\alpha_1 + b_0\alpha_2, \\ &\vdots \\ a_m\beta_n &= b_n\alpha_m. \end{aligned}$$

We view these as a set of homogeneous equations in the $m+n$ unknowns $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n$. Now if A, B have a non-constant common factor then there are non-zero ϕ, ψ satisfying $\psi A = \phi B$ and so the equations have a non-zero solution. It follows that $\mathrm{Res}(A, B) = 0$. (What we have used here is that if $MX = 0$ is a set of equations in matrix notation where M is a square matrix then a non-zero solution exists in $\mathbb{Q}$ if and only if $\det(M) = 0$.) Conversely if $\mathrm{Res}(A, B) = 0$ then the equations have a non-zero solution in $\mathbb{Q}$. But given such a solution we may clear denominators and so obtain a solution in $\mathbb{Z}$. This solution then gives us non-zero $\phi, \psi \in \mathbb{Z}[x]$ such that $\psi A = \phi B$ and it follows from the Claim that A, B have a non-constant common factor. □

It is worth noting that the resultant can be defined for $f, g \in D[x]$ where D is any unique factorization domain. It is clear that the resultant is then an element of D. The preceding theorem also holds in this general case (in the proof we replace $\mathbb{Z}$ with D; in place of $\mathbb{Q}$ we use the so called *field of fractions* of D which is constructed from D in exactly the same way that $\mathbb{Q}$ is constructed from $\mathbb{Z}$; in fact for this construction we only need D to be an integral domain). In this more general situation we sometimes indicate x explicitly by writing $\mathrm{Res}_x(f, g)$.

Lemma 5.4 *Let A, B, p, A_p, B_p be as above (so p does not divide the leading coefcient of one of the polynomials) and put $G = \gcd(A, B)$. Assume that $A_p \neq 0$ and $B_p \neq 0$. If $p \not| \mathrm{Res}(A/G, B/G)$ then $\gcd(A_p, B_p) = G_p$ (recall that this equality is to be interpreted up to non-zero constant multiples of the polynomials).*

Proof A/G, B/G are coprime, $G_p \neq 0$ and

$$\gcd(A_p, B_p) = G_p \gcd(A_p/G_p, B_p/G_p).$$

Now the conclusion of the lemma does not hold if and only if $\gcd(A_p/G_p, B_p/G_p) \neq 1$. Since $A_p \neq 0$ and $B_p \neq 0$ it follows that $A_p/G_p \neq 0$ and $B_p/G_p \neq 0$. If $\gcd(A_p/G_p, B_p/G_p) \neq 1$ then $\mathrm{Res}(A_p/G_p, B_p/G_p) = 0$. But this means that $\mathrm{Res}(A/G, B/G) \equiv 0 \pmod{p}$, in other words $p \mid \mathrm{Res}(A/G, B/G)$ which is contrary to assumption. □

Note that in our case we choose primes p that fail to divide one of the leading coefficients. In general this ensures only that one of the polynomials is non-zero modulo p. However our polynomials are primitive and so no prime divides all of the coefficients of either one and hence neither goes to 0 modulo p. It *is* possible that the degree of one of them drops but the degree of the other will not drop, this is the reason for not insisting that both $a_m \neq 0$ and $b_n \neq 0$ in the definition of the resultant. Note also that $\mathrm{Res}(A/G, B/G) \neq 0$ since A/G, B/G do not have a non-constant common factor (as G is their gcd); Theorem 5.6 now completes the claim. Thus the lemma tells us that there are only finitely many bad primes so that if we keep trying we are bound to find enough good ones. Of course if we knew $\mathrm{Res}(A/G, B/G)$ we could avoid the bad ones altogether but as we do not know G we cannot find the resultant (at least not by direct computation).

Algorithm: $MODGCD(A, B) \mapsto G$

$(A, B \in \mathbb{Z}[x]$ are primitive.)

1. $g := \gcd(\mathrm{lc}(A), \mathrm{lc}(B))$;
$M := 2g\ Landau\_Mignote\_Bound(A, B)$;

2. $p :=$ new prime not dividing g;

3. $C_p := \gcd(A_p, B_p)$; (ensure $\mathrm{lc}(C_p) = 1$)
$G_p := (g \bmod p)C_p$ in $\mathbb{Z}_p[x]$

4. **if** $\deg(G_p) = 0$ **then return** 1 **fi**;
$P := p$;
$G := G_p$;

5. **while** $P \leq M$ **do**
$p :=$ new prime not dividing g;
$C_p := \gcd(A_p, B_p)$; (ensure $\mathrm{lc}(C_p) = 1$)
$G_p := (g \bmod p)C_p$;
if $\deg(G_p) < \deg(G)$ **then goto 4 fi**; (all the previous primes were unlucky)
if $\deg(G_p) = \deg(G)$ **then**
$G := CRA(G, G_p, P, p)$; (we apply CRA_2 to corresponding coefficients of G, G_p)
$P := pP$
fi
od

6. $H := \mathrm{pp}(G)$;
if $H \mid A$ **and** $H \mid B$ **then return** H **fi**;
goto 2 (all the primes were unlucky)

We make some remarks concerning the algorithm.

1. In step 1 we multiply the Landau-Mignotte bound by $g = \gcd(\mathrm{lc}(A), \mathrm{lc}(B))$. This enables us in steps 3 and 4 to normalize $\gcd(A_p, B_p)$ so that it has leading coefficient $(g \bmod p)$.

2. In steps 2 and 3 we require new primes. Every CA system has a large list of primes whose size is about one computer word. Thus in practice we can obtain new primes quickly.

3. Remember that in applying CRA we centre solutions around 0 since we need to recover signed integers (this is also the reason for multiplying the Landau-Mignotte bound by 2 in step 1).

We now look at an example. Let

$$\begin{aligned} A &= (x-2)(x+1)(x^3+2x-1) \\ &= x^5 - x^4 - 3x^2 - 3x + 2, \\ B &= (x-2)^2(x+1)^2 \\ &= x^4 - 2x^3 - 3x^2 + 4x + 4. \end{aligned}$$

This yields

$$\begin{aligned} g &= 1, \\ M &= 2 \cdot 1 \cdot 2^4 \cdot 1 \cdot \min(\sqrt{24}, \sqrt{46}) \\ &\leq 160. \end{aligned}$$

We then have:

Algorithm:

$p = 2:\ G_2 = x^3 + x,$
$\quad P = 2,$
$\quad G = x^3 + x,$

$p = 3:\ G_3 = x^2 - x + 1$, which shows that 2 was unlucky;
$\quad P = 3,$
$\quad G = x^2 - x + 1$

$p = 5:\ G_5 = x^2 - x - 2,$
$\quad G = x^2 - x - 2$, this is $\gcd(A, B)$.

The example shows that it pays to check for $G \mid A$ and $G \mid B$ in the middle of the algorithm as well. However this is a fairly expensive test so we don't necessarily make much gain in speed on the average (a reasonable compromise is to compare the result at each stage with that of the previous stage and if there has been no change then carry out the divisibility test).

5.5 Modular gcd Algorithm for Multivariate Polynomials

Here we consider elements of $\mathbb{Z}[x_1, \ldots, x_n]$. We could try to work in $\mathbb{Q}(x_1, \ldots, x_{n-1})[x_n]$ but then we only get the dependence of the gcd on x_n, the other indeterminates are just units. Alternatively we could use Gauss' lemma:

$$\gcd(A, B) = \gcd(\mathrm{cont}(A), \mathrm{cont}(B)) \gcd(\mathrm{pp}(A), \mathrm{pp}(B)).$$

Here the first gcd is for elements from $\mathbb{Z}[x_1, \ldots, x_{n-1}]$ which may be computed recursively while the second gcd is computed in $\mathbb{Q}(x_1, \ldots, x_{n-1})[x_n]$ which may be computed by Euclid's algorithm. This works but the coefficients increase in size far too much.

We generalize the modular approach by working in $\mathbb{Z}[x_1, \ldots, x_{n-1}][x_n]$ where our coefficients come from $\mathbb{Z}[x_1, \ldots, x_{n-1}]$ and the main indeterminate is x_n. In this method we compute modulo irreducible polynomials in $\mathbb{Z}[x_1, \ldots, x_{n-1}]$. In fact we use linear polynomials of the form $x_s - c$ where $1 \leq s < n$ and $c \in \mathbb{Z}$, these make the computations easier.

Consider now the polynomial ring $R[y][x]$ where R is a UFD (so that $R[y]$ is also a UFD). In our case $R = \mathbb{Z}[x_1, \ldots, x_{n-2}]$, $y = x_{n-1}$ and $x = x_n$. For a polynomial $F \in R[y][x]$ and $r \in R$ we let F_{y-r} stand for $(F \bmod y - r)$. Note that this is the result of substituting r for y in F.

Lemma 5.5 *Let $A, B \in R[y][x]$ and $r \in \mathbb{Z}$. If $y - r$ does not divide both $\mathrm{lc}_x(A)$ and $\mathrm{lc}_x(B)$ then*

$$\deg_x(\gcd(A_{y-r}, B_{y-r})) \geq \deg_x(\gcd(A, B)).$$

Lemma 5.6 *Let A, B, r be as above and $G = \gcd(A, B)$. If $y - r \not| \mathrm{Res}_x(A/G, B/G)$ then $\gcd(A_{y-r}, B_{y-r}) = G_{y-r}$.*

The analogue to the Landau-Mignotte bound is much easier to derive: let C be a factor of A in $R[y][x]$. Then $\deg_y(C) \leq \deg_y(A)$; this follows from the fact that R is a UFD and so has no zero-divisors.

Algorithm: *MODGCDm*$(A, B, n, s) \mapsto C$;

(n is the number of variables and s is the index of the variable being eliminated.)

1. **if** $s = 0$ **then** $C :=$ *univariate_gcd*(A, B); **return** C **fi**;

2. $M := 1 + \min(\deg_{x_s}(A), \deg_{x_s}(B))$;

3. $r :=$ an integer s.t. $\deg_{x_n}(A_{x_s - r}) = \deg_{x_n}(A)$ or $\deg_{x_n}(B_{x_s - r}) = \deg_{x_n}(B)$;
$C_r := MODGCDm(A_{x_s - r}, B_{x_s - r}, n, s - 1)$;

4. $R := x_s - r$;
$m := 1$;
$C := C_r$;

5. **while** $m \leq M$ **do**
$r :=$ a new integer s.t. $\deg_{x_n}(A_{x_s - r}) = \deg_{x_n}(A)$ or $\deg_{x_n}(B_{x_s - r}) = \deg_{x_n}(B)$;
$C_r := MODGCDm(A_{x_s - r}, B_{x_s - r}, n, s - 1)$;
if $\deg_{x_n}(C_r) < \deg_{x_n}(C)$ **then goto 3 fi**;
if $\deg_{x_n}(C_r) = \deg_{x_n}(C)$ **then**
$C := CRA(C, C_r, R, x_s - r)$;
$R := (x_s - r)R$;
$m := m + 1$
fi
od;

6. **if** $C \mid A$ **and** $C \mid B$ **then return** C **fi**;
goto 2

We look at an example in $\mathbb{Z}[x, y]$. Let

$$A = (2xy - y + x^2)(xy^2 + x^3 - 3),$$
$$B = (2xy - y + x^2)(y^2 - xy + 2).$$

We have

$$M = 1 + \min(\deg_x(A), \deg_x(B)) = 4.$$

The algorithm proceeds as follows:

$r = 1$: $\gcd(A_{x-1}, B_{x-1}) = y + 1$.

$r = 2$: $\gcd(A_{x-2}, B_{x-2}) = 3y + 4$. Now we use the Chinese Remainder Theorem for polynomials to obtain

$$C = (2x - 1)y + (3x - 2).$$

$r = 3$: $\gcd(A_{x-3}, B_{x-3}) = 5y + 9$. This time the CRT yields

$$C = (2x - 1)y + x^2$$

and this is the gcd (the algorithm would actually take another step).

Unfortunately there is a problem with the algorithm: we have to use the CRT in $\mathbb{Z}[x_1, \ldots, x_{n-1}]$ but this is not a Euclidean domain. The problem we must solve is: given $p_1(x_1, \ldots, x_s), p_2(x_1, \ldots, x_s)$ and moduli r_1, r_2 find $p(x_1, \ldots, x_s)$ (all in $\mathbb{Z}[x_1, \ldots, x_s]$) such that

$$p \equiv p_1 \pmod{r_1},$$
$$p \equiv p_2 \pmod{r_2},$$

such that $\deg_{x_s}(p) < \deg_{x_s}(r_1) + \deg_{x_s}(r_2)$. We can solve the problem by using the embedding

$$\mathbb{Z}[x_1, \ldots, x_s] \subseteq \mathbb{Q}(x_1, \ldots, x_{s-1})[x_s].$$

Now the right hand side is a Euclidean domain so we solve our problem here to obtain a unique solution p such that $\deg_{x_s}(p) < \deg_{x_s}(r_1) + \deg_{x_s}(r_2)$. Now if $p \in \mathbb{Z}[x_1, \ldots, x_s]$ we are done and otherwise the irreducible polynomials were unlucky so we have to start again.

The modular approach is the fastest currently known. Most systems start with some heuristics. We note also that all systems use a recursive representation of polynomials for this algorithm.