Factoring multivariate integral polynomials

A.K. Lenstra
Mathematisch Centrum
Kruislaan 413
1098 SJ Amsterdam
The Netherlands

Abstract.

We present an algorithm to factor polynomials in several variables with integral coefficients that is polynomial-time in the degrees of the polynomial to be factored. Our algorithm generalizes the algorithm presented in [7] to factor integral polynomials in one variable.

1. Introduction.

The problem of factoring polynomials with integral coefficients remained open for a long time, i.e. no polynomial-time factoring algorithm was known. The best known algorithms took exponential-time in the worst case; these algorithms had to consider a possibly exponential number of combinations of p-adic factors before the true factors could be found or irreducibility could be decided. In [1] it was proven that the problem of factorization in $\mathbb{Z}[X]$ belongs to $NP \cap co\text{-}NP$, which made its membership of P quite likely [2]. That this was indeed the case, was proven in [7] where a polynomial-time algorithm for factoring in $\mathbb{Z}[X]$ was given. This algorithm is based on the following three observations:

(1.1) The multiples of degree $< m$ of a p-adic factor together form a lattice in $\mathbb{Z}^m$;

(1.2) If this p-adic factor is computed up to a high enough precision, then the factor we are looking for is the shortest vector in this lattice;

(1.3) An approximation of the shortest vector in such a lattice can be found in polynomial-time by means of the so-called *basis reduction algorithm*.

In this paper we show that (1.1) and (1.2) can be generalized to polynomials in $\mathbb{Z}[X_1, X_2, \ldots, X_t]$ in an elementary way, for any $t \geq 2$. Combined with the same basis reduction algorithm as in (1.3), this leads to a polynomial-time algorithm for factoring in $\mathbb{Z}[X_1, X_2, \ldots, X_t]$. In [8, 9, 10] we show that the above three points can be applied to various other kinds of polynomial factoring problems as well (like multivariate polynomials over finite fields or over algebraic number fields). Another approach to multivariate integral polynomial factorization is given in [5]. There the multivariate case is first reduced in polynomial-time to the bivariate case, next bivariate is reduced to univariate.

For practical purposes we do not recommend any of these polynomial-time algorithms; their running time will be dominated by the rather slow basis reduction algorithm. For polynomials in $\mathbb{Z}[X_1, X_2, \ldots, X_t]$ the algorithm from [12] for instance is very useful, although it is exponential-time in the worst case.

We restrict ourselves in this paper to integral polynomials in two variables; the multivariate case follows immediately from this. In Section 2 we present an important result from [7: Section 1] concerning the basis reduction algorithm mentioned in (1.3). The generalizations of (1.1) and (1.2) to polynomials in $\mathbb{Z}[X,Y]$ are described in Section 3, and in Section 4 we give an outline of the factoring algorithm, and we analyze its running time.

2. The basis reduction algorithm.

The basis reduction algorithm from [7: Section 1] makes it possible to determine in polynomial-time a reasonable approximation of the shortest vector in a lattice. We will not give a description of the algorithm here. It will suffice to summarize those results from [7: Section 1] that we will need here.

Let $b_1, b_2, \ldots, b_n \in \mathbb{Z}^n$ be linearly independent. For our purposes we may assume that the $n\times n$ matrix having $b_1, b_2, \ldots, b_n$ as columns is upper-triangular. The i-dimensional lattice $L_i \subset \mathbb{Z}^i$ with basis $b_1, b_2, \ldots, b_i$ is defined as $L_i = \Sigma_{j=1}^{i} \mathbb{Z} b_j = \{\Sigma_{j=1}^{i} r_j b_j : r_j \in \mathbb{Z}\}$. We put $L = L_n$.

(2.1) Proposition. (cf. [7: (1.11), (1.26), (1.37)]) Let $B \in \mathbb{Z}_{\geq 2}$ be such that $|b_j|^2 \leq B$ for $1 \leq j \leq n$, where $|\;|$ denotes the ordinary Euclidean length. The basis reduction algorithm as described in [7: (1.15)] determines a vector $\tilde{b} \in L$ such that $\tilde{b}$ belongs to a basis for L, and such that $|\tilde{b}|^2 \leq 2^{n-1}|x|^2$ for every $x \in L$, $x \neq 0$; the algorithm takes $O(n^4 \log B)$ elementary operations on integers having binary length $O(n \log B)$. Furthermore, during the first $O(i^4 \log B)$ operations (on integers having binary length $O(i \log B)$), vectors $\tilde{b}_i \in L_i$, belonging to a basis for L_i, are determined such that $|\tilde{b}_i|^2 \le 2^{i-1}|x_i|^2$ for every $x_i \in L_i$, $x_i \ne 0$, for $1 \le i \le n$. □

So, we can find a reasonable approximation of the shortest vector in L in polynomial-time. But also we find, during this computation, approximations of the shortest vectors of the lattices L_i without any time loss.

3. Factors and lattices.

We describe how to generalize (1.1) and (1.2) to polynomials in $\mathbb{Z}[X,Y]$. Let $f \in \mathbb{Z}[X,Y]$ be the polynomial to be factored; we may assume that f has no multiple factors, i.e. f is *square-free*. Furthermore we assume that f is *primitive* with respect to X, i.e. the greatest common divisor of the coefficients in $\mathbb{Z}[Y]$ of f equals one. We denote by $\delta_X f$ and $\delta_Y f$ the degrees of f in X and Y respectively, and by $\ell c(f)$ the *leading coefficient* of f with respect to X. We put $n_X = \delta_X f$ and $n_Y = \delta_Y f$.

Suppose that we are given a prime number p, an integer s and a positive integer k. By (s_1) we denote the ideal generated by p and (Y-s), and by (s_k) we denote the ideal generated by p^k and $(Y-s)^{n_Y+1}$. In Section 4 we will see how to find a polynomial $h \in \mathbb{Z}[X,Y]$ such that:

(3.1) $\ell c(h) = 1$,

(3.2) $(h \bmod (s_k))$ divides $(f \bmod (s_k))$ in $\mathbb{Z}[X,Y]/(s_k)$,

(3.3) $(h \bmod (s_1)) \in (\mathbb{Z}/p\mathbb{Z})[X]$ is irreducible in $(\mathbb{Z}/p\mathbb{Z})[X]$,

(3.4) $(h \bmod (s_1))^2$ does not divide $(f \bmod (s_1))$ in $(\mathbb{Z}/p\mathbb{Z})[X]$.

We put $\ell = \delta_X h$; so $0 < \ell \le n_X$.

Let $h_0 \in \mathbb{Z}[X,Y]$ be the irreducible factor of f for which $(h \bmod (s_1))$ divides $(h_0 \bmod (s_1))$ in $(\mathbb{Z}/p\mathbb{Z})[X]$ (or equivalently $(h \bmod (s_k))$ divides $(h_0 \bmod (s_k))$ in $\mathbb{Z}[X,Y]/(s_k)$, cf. [7: (2.5)]); notice that h_0 is unique up to sign.

(3.5) Let m_X and m_Y be two integers with $\ell \le m_X < n_X$ and $0 \le m_Y \le \delta_Y \ell c(f)$. We define L as the collection of polynomials $g \in \mathbb{Z}[X,Y]$ such that

(i) $\delta_X g \le m_X$,

(ii) $\delta_Y g \le n_Y$,

(iii) $\delta_Y \ell c(g) \le m_Y$,

(iv) $(h \bmod (s_k))$ divides $(g \bmod (s_k))$ in $\mathbb{Z}[X,Y]/(s_k)$.

Putting $M = m_X(n_Y+1)+m_Y+1$ it is not difficult to see that L is an M-dimensional lattice contained in $\mathbb{Z}^M$, where we identify polynomials in L and M-dimensional vectors in the usual way (i.e. $\Sigma_{i=0}^{m_X-1} \Sigma_{j=0}^{n_Y} a_{ij} X^i Y^j + \Sigma_{j=0}^{m_Y} a_{m_X j} X^{m_X} Y^j$ is identified with $(a_{00}, a_{01}, \ldots, a_{0n_Y}, a_{10}, \ldots, a_{m_X-1\, n_Y}, a_{m_X 0}, \ldots, a_{m_X m_Y})$). Because of (3.1) a basis for L is given by

$$\{p^k Y^j X^i:\ 0 \le j \le n_Y,\ 0 \le i < \ell\} \cup$$
$$\{(hY^j \bmod (s_k))X^{i-\ell}:\ (0 \le j \le n_Y \text{ and } \ell \le i < m_X) \text{ or } (0 \le j \le m_Y \text{ and } i = m_X)\}.$$

This generalizes (1.1) (cf. [7: (2.6)]). We now come to (1.2). The *height* g_{max} of a polynomial g is defined as the maximal absolute value of any of its integral coefficients. We prove that, if k and s are suitably chosen, then a vector of small height in L must lead to a factorization of f.

<u>(3.6) Proposition</u>. Suppose that $g \in L$ satisfies

$$(3.7)\qquad |s|^{n_Y+1} > (e^{n_X+n_Y} f_{max}\sqrt{(n_X+1)(n_Y+1)})^{m_X}(g_{max}\sqrt{(m_X+1)(n_Y+1)})^{n_X}$$

and

$$(3.8)\qquad p^k > (e^{n_X+n_Y} f_{max}\sqrt{(n_X+1)(n_Y+1)})^{m_X}(g_{max}\sqrt{(m_X+1)(n_Y+1)})^{n_X}(1+(1+|s|)^{n_Y+1})^{n_Y(n_X+m_X-1)}.$$

Then h_0 divides g in $\mathbb{Z}[X,Y]$, and in particular $\gcd(f,g) \ne 1$.

<u>Proof</u>. Suppose that $\gcd(f,g) = 1$. This implies that the resultant $R \in \mathbb{Z}[Y]$ of f and g is unequal to zero. Using the result from [4] one proves that

$$(3.9)\qquad |R| < (f_{max}\sqrt{(n_X+1)(n_Y+1)})^{m_X}(g_{max}\sqrt{(m_X+1)(n_Y+1)})^{n_X},$$

where $|R|$ denotes the ordinary Euclidean length of the vector identified with R. Since $(h \bmod (s_k))$ divides both $(f \bmod (s_k))$ and $(g \bmod (s_k))$, the polynomials f and g have a non-trivial common divisor in $\mathbb{Z}[X,Y]/(s_k)$, so that R must be zero modulo the ideal generated by p^k and $(Y-s)^{n_Y+1}$. The polynomial $(Y-s)^{n_Y+1}$ cannot divide R, because this would imply, according to [11: Theorem 1], that $|s|^{n_Y+1} \le |R|$, which is, combined with (3.9), a contradiction with (3.7). Therefore $(R \bmod (Y-s)^{n_Y+1})$ has to be zero modulo p^k. Using induction on n_Y+1 it is easy to prove that

$$(R \bmod (Y-s)^{n_Y+1})_{max} \le R_{max}(1+(1+|s|)^{n_Y+1})^{n_Y(n_X+m_X-1)},$$

so that, with $R_{max} \le |R|$ and (3.8), it follows that $(R \bmod (Y-s)^{n_Y+1})$ cannot be zero modulo p^k. We conclude that $\gcd(f,g) \neq 1$.

Suppose that h_0 does not divide g. So h_0 does not divide $r = \gcd(f,g)$, so $(h \bmod (s_k))$ divides $((f/r) \bmod (s_k))$. Because f/r divides f, we find from [3] that $(f/r)_{max} \leq e^{n_X+n_Y} f_{max}$. This implies that the above reasoning applies to f/r and the same polynomial g in L, so that $\gcd(f/r,g) \neq 1$. This is a contradiction with $r = \gcd(f,g)$, because f is square-free. □

(3.10) Proposition. Suppose that s and k are chosen in such a way that (3.7) and (3.8) are satisfied with g_{max} replaced by $2^{(M-1)/2}\sqrt{M}\, e^{n_X+n_Y} f_{max}$. Let $\tilde{b}$ be as in (2.1) the result of an application of the basis reduction algorithm to the M-dimensional lattice L as defined in (3.5). Then $h_0 \in L$ if and only if (3.7) and (3.8) are satisfied with g replaced by $\tilde{b}$.

Proof. To prove the "if"-part, assume that (3.7) and (3.8) hold with g_{max} replaced by $\tilde{b}_{max}$. According to (3.6) this implies that h_0 divides $\tilde{b}$, so that $h_0 \in L$.

To prove the "only if"-part, assume that $h_0 \in L$. Because h_0 divides f, we find from [3] that $(h_0)_{max} \leq e^{n_X+n_Y} f_{max}$. So there exists a non-zero vector in L with Euclidean length bounded by $\sqrt{M}\, e^{n_X+n_Y} f_{max}$. Application of (2.1) yields that $\tilde{b}_{max} \leq |b| \leq 2^{(M-1)/2}\sqrt{M}\; e^{n_X+n_Y} f_{max}$. Combined with the above choices of s and k, this implies that (3.7) and (3.8) hold with g replaced by $\tilde{b}$. □

4. Description of the algorithm.

In this section we present the polynomial-time algorithm to factor f. First we give an algorithm to determine the factor h_0, given p, s and h. After that, we will see how p and s have to be chosen.

(4.1) Let p, s and h be as in Section 3, such that (3.1), (3.3), (3.4) and (3.2) with k replaced by 1 are satisfied. Assume that s satisfies the condition in (3.10) with m_X and m_Y replaced by n_X-1 and $\delta_Y \ell c(f)$ respectively:

$$(4.2) \qquad |s|^{n_Y+1} > \left(e^{n_X+n_Y} f_{max}\sqrt{(n_X+1)(n_Y+1)}\right)^{n_X-1}\left(2^{(M-1)/2}\sqrt{M}\, e^{n_X+n_Y} f_{max}\sqrt{n_X(n_Y+1)}\right)^{n_X}$$

where $M = (n_X-1)(n_Y+1)+\delta_Y \ell c(f)+1$. We describe an algorithm that determines h_0, the irreducible factor of f such that $(h \bmod (s_1))$ divides $(h_0 \bmod (s_1))$ in $(\mathbb{Z}/p\mathbb{Z})[X]$.

We may assume that $\ell = \delta_X h < n_X$. Take k minimal such that the condition from (3.10) is satisfied with m_X and m_Y replaced by n_X-1 and $\delta_Y \ell c(f)$ respectively:

$$(4.3) \qquad p^k > (e^{n_X+n_Y} f_{max} \sqrt{(n_X+1)(n_Y+1)})^{n_X-1} (2^{(M-1)/2} \sqrt{M}\, e^{n_X+n_Y} f_{max} \sqrt{n_X(n_Y+1)})^{n_X} \cdot (1+(1+|s|)^{n_Y+1})^{2n_Y(n_X-1)}.$$

Next modify h in such a way that (3.2) also holds for this value of k; because of (3.4) this can be done by means of Hensel's lemma [13].

Apply Proposition (2.1) to the M-dimensional lattice L as defined in (3.5) for each of the values of $M = \ell(n_Y+1)+1, \ell(n_Y+1)+2, \ldots, \ell(n_Y+1)+\delta_Y \ell c(f)+1, (\ell+1)(n_Y+1)+1, \ldots, (n_X-1)(n_Y+1)+\delta_Y \ell c(f)+1$ in succession (so, for $m_X = \ell, \ell+1, \ldots, n_X-1$ in succession and for every value of m_X the values $m_Y = 0, 1, \ldots, \delta_Y \ell c(f)$ in succession). But stop as soon as a vector $\tilde{b}$ is found satisfying (3.7) and (3.8) with g replaced by $\tilde{b}$.

If such a vector $\tilde{b}$ is found for a certain value of M ($m_X = m_{X0}$ and $m_Y = m_{Y0}$), then we know from (3.10) that $h_0 \in L$. Since we try the values of M in succession this implies that $\delta_X h_0 = m_{X0}$ and $\delta_Y \ell c(h_0) = m_{Y0}$. By (3.6) h_0 divides $\tilde{b}$, so that $\delta_X \tilde{b} = m_{X0}$ and $\delta_Y \ell c(\tilde{b}) = m_{Y0}$. So $\tilde{b} = ch_0$ for some $c \in \mathbb{Z}$, but $h_0 \in L$ and $\tilde{b}$ belongs to a basis for L, so $\tilde{b} = \pm h_0$.

If no such vector $\tilde{b}$ was found, then (3.10) implies that $\delta_X h_0 > n_X-1$, so that $h_0 = f$, because f is primitive.

This finishes the description of Algorithm (4.1).

<u>(4.4) Proposition</u>. Denote by $m_{X0} = \delta_X h_0$ the degree in X of the irreducible factor h_0 of f that is found by Algorithm (4.1). Then the number of arithmetic operations needed by Algorithm (4.1) is $O(m_{X0}(n_X^5 n_Y^5 + n_X^4 n_Y^4 \log(f_{max}) + n_X^4 n_Y^6 \log(|s|) + n_X^3 n_Y^4 \log p))$ and the integers on which these operations have to be performed each have binary length $O(n_X^3 n_Y^2 + n_X^2 n_Y \log(f_{max}) + n_X^2 n_Y^3 \log(|s|) + n_X n_Y \log p)$.

<u>Proof</u>. Let M_1 be the largest value of M for which (2.1) is applied; so $M_1 = O(m_{X0} n_Y)$. It follows from (2.1) that the number of operations needed for the applications of the basis reduction algorithm for $\ell(n_Y+1)+1 \leq M \leq M_1$ is equal to the number of operations needed for $M = M_1$ only. Assuming that the coefficients of the initial basis for L are reduced modulo p^k, we find, using (4.3), that the following holds for the bound

B on the length of these vectors:

$$\log B = O(n_X^2 n_Y + n_X \log(f_{max}) + n_X n_Y^2 \log(|s|) + \log p).$$

With $M_1 = O(m_{X0} n_Y)$ and (2.1) this gives the estimates in (4.4).

The verification that the same estimates are valid for the application of Hensel's lemma is straightforward [13]. □

We now describe how s and p have to be chosen. First, s must be chosen such that $(f \bmod (Y-s)) = f(X,s)$ remains square-free, and such that (4.2) holds. The resultant R of f and its derivative f' with respect to X is a non-zero polynomial in $\mathbb{Z}[Y]$ of degree $\leq n_Y(2n_X-1)$. Therefore we can find in $O(n_X n_Y)$ trials the minimal integer s such that s is not a zero of R, and such that (4.2) holds. It is easily verified that $\log(|s|) = O(n_X^2 + n_X \log(f_{max}))$.

Next we choose p as the smallest prime number not dividing the resultant of $f(X,s)$ and $f'(X,s)$. Since $\log(f(X,s)_{max}) = O(n_X^2 n_Y + n_X n_Y \log(f_{max}))$, it follows as in the proof of [7: (3.6)] that $p = O(n_X^3 n_Y + n_X^2 n_Y \log(f_{max}))$.

The complete factorization of $(f \bmod (s_1))$ can be determined by means of Berlekamp's algorithm [6: section 4.6.2]; notice that (3.4) holds for every factor $(h \bmod (s_1))$ of $(f \bmod (s_1))$, because of the choice of p, and that this factorization can be found in polynomial-time, because of the bound on p. The algorithm to factor f completely now follows by repeated application of Algorithm (4.1). The above bounds on $\log(|s|)$ and p, combined with (4.4) and the fact that a factor g of f satisfies $\log(g_{max}) = O(n_X + n_Y + \log(f_{max}))$ (cf. [3]), yields the following theorem.

(4.5) Theorem. The number of arithmetic operations needed to factor f completely is $O(n_X^7 n_Y^6 + n_X^6 n_Y^6 \log(f_{max}))$, and the integers on which these operations have to be performed each have binary length $O(n_X^4 n_Y^3 + n_X^3 n_Y^3 \log(f_{max}))$. □

5. Conclusion.

We have shown that basically the same ideas that were used for the polynomial-time algorithm for factoring in $\mathbb{Z}[X]$ lead to a polynomial-time factoring algorithm in $\mathbb{Z}[X,Y]$ (Theorem (4.5)). Our method can be generalized to polynomials in $\mathbb{Z}[X_1, X_2,$

$\ldots, X_t]$. The evaluation $(Y=s)$ is then replaced by $(X_2=s_2, X_3=s_3, \ldots, X_t=s_t)$, where the integers s_i have to satisfy conditions similar to (4.2). It will not be surprising that in this case the estimates become rather complicated.

A somewhat simpler algorithm results if we use the algorithm from [7]; the details of this algorithm, which is similar to the one described in this paper, can be found in [10].

References.

1. D.G. Cantor, Irreducible polynomials with integral coefficients have succinct certificates, J. of Algorithms 2 (1981), 385-392.

2. M.R. Garey, D.S. Johnson, Computers and intractability, Freeman, San Francisco 1979.

3. A.O. Gel'fond, Transcendental and algebraic numbers, Dover Publ., New York 1960.

4. A.J. Goldstein, R.L. Graham, A Hadamard-type bound on the coefficients of a determinant of polynomials, SIAM Rev. 16 (1974), 394-395.

5. E. Kaltofen, On the complexity of factoring polynomials with integer coefficients, Ph.D. thesis, Rensselaer Polytechnic Institute, August 1982.

6. D.E. Knuth, The art of computer programming, vol. 2, Seminumerical algorithms, Addison Wesley, Reading, second edition 1981.

7. A.K. Lenstra, H.W. Lenstra, Jr., L. Lovász, Factoring polynomials with rational coefficients, Math. Ann. 261 (1982), 515-534.

8. A.K. Lenstra, Factoring polynomials over algebraic number fields, Report IW 213/82, Mathematisch Centrum, Amsterdam 1982 (also Proceedings Eurocal 83).

9. A.K. Lenstra, Factoring multivariate polynomials over finite fields, Report IW 221/83, Mathematisch Centrum, Amsterdam 1983 (also Proceedings 15th STOC).

10. A.K. Lenstra, to appear.

11. M. Mignotte, An inequality about factors of polynomials, Math. Comp. 28 (1974), 1153-1157

12. P.S. Wang, An improved multivariate polynomial factoring algorithm, Math. Comp. 32 (1978), 1215-1231.

13. D.Y.Y. Yun, The Hensel lemma in algebraic manipulation, MIT, Cambridge 1974; reprint: Garland publ. Co., New York 1980.