A Recent Algorithm for the Factorization of Polynomials

Arjen K. Lenstra
Department of Computer Science
The University of Chicago, Ryerson Hall
1100 E. 58th Street, Chicago, IL 60637, USA

1. INTRODUCTION

The last few years a lot of attention has been paid to the problem of factoring polynomials with rational coefficients. An important result was the discovery of a *polynomial-time* factoring algorithm [7]. The purpose of this note is to provide an informal description of this new algorithm.

It is well known that a polynomial in $\mathbb{Q}[X]$ can be decomposed into irreducible factors in $\mathbb{Q}[X]$ and that this factorization is unique up to units. Such a factorization is equivalent to the factorization of a *primitive* polynomial with integral coefficients, where a polynomial is called primitive if the greatest common divisor of its coefficients equals 1. Throughout this note we will therefore restrict ourselves to primitive integral polynomials.

In VAN DER WAERDEN [13] it is shown that the factorization of a polynomial in $\mathbb{Z}[X]$ is effectively computable. The method described there was invented in 1793 by the German astronomer VON SCHUBERT, and later re-invented by KRONECKER; it is usually referred to as *Kronecker's method*. For practical purposes this algorithm can hardly be recommended. A better algorithm was published in 1969 by ZASSENHAUS [15]. It is based on a combination of Berlekamp's algorithm for the factorization of polynomials over finite fields [6, Section 4.6.2] and Hensel's lemma [6, Exercise 4.6.2.22], and is therefore called the *Berlekamp-Hensel algorithm*. Zassenhaus' method performs quite well in practice, and there is some evidence that its expected running time is a polynomial function of the degree of the polynomial to be factored [2]. It has however one important disadvantage: its worst-case running time is an exponential function of the degree. Polynomials that exhibit the exponential behaviour of the Berlekamp-Hensel algorithm can easily be constructed [5].

In 1982 an algorithm was presented whose running time, when applied to some polynomial f in $\mathbb{Z}[X]$, is always bounded by a fixed polynomial function of the degree and the coefficient-size of f [7]. A simplified and slightly improved version of this algorithm was given in [4] and [12]. This latter version, which we will follow here, is based on the following observation. The irreducible factors in $\mathbb{Z}[X]$ of f can be regarded as the minimal polynomials (in $\mathbb{Z}[X]$) of its roots. Therefore, to find an irreducible factor of f, it suffices to determine the minimal polynomial of one of its roots. The minimal polynomial of a root α of f immediately follows from an integral linear combination of minimal degree among the powers of α. In Section 2 it is shown that the problem of finding such a relation among the powers of α can be reduced to the problem of finding a relatively short vector in a certain subset of a real vector space. Such a short vector can then be found by means of the *basis reduction algorithm* , as is explained in Section 3.

2. REDUCTION TO FINDING SHORT VECTORS

Let f in $\mathbb{Z}[X]$ be the polynomial to be factored and let α be one of its roots. For simplicity we assume that α is real; the general case easily follows from this. Denote by h in $\mathbb{Z}[X]$ the minimal polynomial of α. Obviously, this polynomial h is an irreducible factor of f.

Suppose the degree of h equals m, for some positive integer m. Let c be some fixed positive integer. Below we will show how this integer should be chosen. For an arbitrary polynomial $g \in \mathbb{Z}[X]$ of degree at most m we denote by $\bar{g}$ the $(m+2)$-dimensional vector having the coefficient of X^{i-1} of g as ith coordinate, for $0<i \leqslant m+1$, and with last coordinate $c \cdot g(\alpha)$. By L_m we denote the subset of $\mathbb{R}^{m+2}$ consisting of these vectors $\bar{g}$; notice that the $(m+2)$-dimensional vector $\bar{h}$ is contained in L_m. There is a natural correspondence between the vectors $\bar{g}$ and integral linear combinations of degree at most m among the powers of α: the first $m+1$ coordinates of $\bar{g}$ correspond to the coefficients of the integral linear combination, and the last coordinate of $\bar{g}$ is the value of that particular combination, multiplied by c. In this Section we show that a relatively short non-zero vector in L_m leads to the coefficients of h, where we use the ordinary Euclidean norm in $\mathbb{R}^{m+2}$ (denoted $|\ |$).

Because h is a factor of f, there exists an upper bound on the absolute value of the coefficients of h that depends only on f [9]. Combined with $h(\alpha)=0$, we find that there is a bound $B_f \geqslant 2$, only depending on f and not on c, such that $|\bar{h}| \leqslant B_f$. We claim that for any $C>1$ the value for c can be chosen such that $|\bar{g}| > C \cdot B_f$ if $\gcd(h,g)=1$. This means that we can choose c in such a way that any non-zero vector $\bar{g}$ that is not much longer than $\bar{h}$, leads to h. Namely, if $|\bar{g}| \leqslant C \cdot B_f$ then $\gcd(h,g) \neq 1$, so that g is an integral multiple of h because h is irreducible and because the degree of g is at most m. Thus h can be found if we can find a vector $\bar{g}$ that is relatively short, i.e., $|\bar{g}| \leqslant C \cdot B_f$ for some $C>1$.

To prove our claim, let $C>1$ be a real number, and let $g \in \mathbb{Z}[X]$ of degree at most m be such that $\gcd(h, g) = 1$. We prove that c can be chosen such that $|\bar{g}| > C\cdot B_f$. Obviously, if the Euclidean length of the vector g (i.e., the vector consisting of the first $m+1$ coordinates of $\bar{g}$) is $> C\cdot B_f$; then also $|\bar{g}| > C\cdot B_f$. Therefore we may assume that the Euclidean length of the vector g is bounded by $C\cdot B_f$; it suffices to prove that c can be chosen such that $|c\cdot g(\alpha)| > C\cdot B_f$.

Denote by n the degree of g. Define the $(m+n)\times(m+n)$ matrix M as the matrix having ith column $X^{i-1}\cdot h$ for $1 \leqslant i \leqslant n$ and $X^{i-n-1}\cdot g$ for $n+1 \leqslant i \leqslant m+n$, where $X^{i-1}\cdot h$ and $X^{i-n-1}\cdot g$ are regarded as $(m+n)$-dimensional vectors. By R we denote the absolute value of the determinant of M, the so-called *resultant* of h and g.

We prove that this resultant R is non-zero. Suppose on the contrary that the determinant of M is zero. This would imply that a linear combination of the columns of M is zero, so that there exist polynomials $a, b \in \mathbb{Z}[X]$ with $\text{degree}(a) < n$ and $\text{degree}(b) < m$ such that $a\cdot h + b\cdot g = 0$. Because $\gcd(h, g) = 1$, we have that h divides b, so that with $\text{degree}(b) < m$, we find $b = 0$, and also $a = 0$. This proves that the columns of M are linearly independent, so that $R \neq 0$. Because the entries of M are integral we even have $R \geqslant 1$.

We add, for $2 \leqslant i \leqslant m+n$, the ith row of M times T^{i-1} to the first row of M, for some indeterminate T. The first row of M then becomes $(h(T), T\cdot h(T), ..., T^{n-1}\cdot h(T), g(T), T\cdot g(T), ..., T^{m-1}\cdot g(T))$. Expanding the determinant of M with respect to the first row, we find that

$$R = |h(T)\cdot(a_0 + a_1\cdot T + ... + a_{n-1}\cdot T^{n-1}) \\ + g(T)\cdot(b_0 + b_1\cdot T + ... + b_{m-1}\cdot T^{m-1})|,$$

where the a_i and b_j are determinants of $(m+n-1)\times(m+n-1)$ submatrices of M. Evaluating the above identity for $T = \alpha$ yields

$$R = |g(\alpha)| \cdot |b_0 + b_1\cdot\alpha + ... + b_{m-1}\cdot\alpha^{m-1}|,$$

because $h(\alpha) = 0$. From $|\bar{h}| \leqslant B_f$, $|g| \leqslant C\cdot B_f$, and Hadamard's inequality it follows that $|b_j| \leqslant (C\cdot B_f)^{m+n-1}$. Because B_f is also an upper bound for the roots of f we get

$$R \leqslant |g(\alpha)| \cdot (C\cdot B_f)^{2m+n-1},$$

so that, with $R \geqslant 1$, we find

$$|g(\alpha)| \geqslant (C\cdot B_f)^{-2m-n+1}.$$

Therefore, in order to get $|c\cdot g(\alpha)| > C\cdot B_f$, it suffices to take $c > (C\cdot B_f)^{3m}$. This proves our claim.

Of course, the degree m of h is not known beforehand. The way in which we apply the above to determine h is as follows.

For some $C > 1$, to be specified in the next section, we take c minimal such that $c > (C\cdot B_f)^{3\cdot\text{degree}(f)}$. Next for $m = 1, 2, ..., \text{degree}(f) - 1$ in succession we do the following. Consider the set L_m of $(m+2)$-dimensional vectors $\bar{g}$ as defined above. Because $(C \cdot B_f)^{3 \cdot \text{degree}(f)} \geq (C \cdot B_f)^{3 \cdot \text{degree}(h)}$, a non-zero vector $\bar{g}$ in L_m satisfying $|\bar{g}| \leq C \cdot B_f$ leads to a polynomial g that has a non-trivial greatest common divisor with h. Therefore, for values of m smaller than the degree of h all non-zero vectors in L_m must have length $> C \cdot B_f$, and there can only be non-zero vectors $\bar{g}$ in L_m satisfying $|\bar{g}| \leq C \cdot B_f$ if m is at least equal to the degree of h, i.e., if $\bar{h}$ is also contained in L_m. And, as reasoned above, if m equals the degree of h, then a reasonably short non-zero vector $\bar{g}$ leads to a polynomial g that is a non-trivial multiple of h. This implies that for $m = \text{degree}(h)$ vector $\bar{h}$ is a shortest non-zero vector in the set L_m, and that $\bar{h}$ can be determined if we can find a non-zero vector in L_m that is longer than $\bar{h}$ by at most a factor C. In the next section we will see that, for some value of $C > 1$, we can always find a non-zero vector in L_m that is at most a factor C longer than a shortest non-zero vector in L_m. Thus the algorithm can be terminated as soon as we succeed in finding a non-zero vector $\bar{g}$ of length at most $C \cdot B_f$. If no such vector is found, then all values for m are smaller than degree(h), so that $h = f$.

REMARK. If α is irrational, then in practice it is impossible to work with an exact representation of α. However, it is not difficult to see that the same arguments as above apply if we use a sufficiently close approximation $\tilde{\alpha}$ to α. It appears that it suffices to have $|\alpha - \tilde{\alpha}| < 2^{-s}$, where s is bounded by a polynomial function of the degree of f and of $\log|f|$. Such an approximation of a root of f can be found in polynomial time, as is shown in [11].

If α is a non-real complex number, then we modify the definition of $\bar{g}$ as follows: for arbitrary $g \in \mathbf{Z}[X]$ of degree at most m we denote by $\bar{g}$ the $(m+3)$-dimensional vector having the coefficient of X^{i-1} of g as ith coordinate, for $0 < i \leq m+1$, and with last two coordinates $c \cdot \text{Re}(g(\alpha))$ and $c \cdot \text{Im}(g(\alpha))$.

3. HOW TO FIND THE SHORTEST VECTOR

In the previous section we have reduced the problem of factoring polynomials with rational coefficients to the problem of finding a relatively short vector in a certain subset L_m of $\mathbf{R}^{m+2}$. Such a subset of a real vector space is usually called a *lattice*. In this section we will discuss the problem of finding short non-zero vectors in a lattice, and we will see that the shortest vector problem from Section 2 can be solved by means of L. Lovász' *basis reduction algorithm*.

Let n and k be positive integers, and let $b_1, b_2, ..., b_k$ be linearly independent vectors in $\mathbf{R}^n$. The *lattice of dimension* k generated by $b_1, b_2, ..., b_k$ is defined as the set

$$\{\sum_{i=1}^{n} r_i b_i : r_i \in \mathbf{Z}\}.$$

The lattice is denoted $L = L(b_1, b_2, ..., b_k)$ and $b_1, b_2, ..., b_k$ is said to be a *basis* for the lattice. Clearly, the set L_m from Section 2 is an $(m+1)$-dimensional lattice generated by $\bar{g}_0, \bar{g}_1, \ldots, \bar{g}_m$ where $g = X^i$, for $i = 0, 1, \ldots, m$.

The shortest vector problem for a lattice $L = L(b_1, b_2, \ldots, b_k)$ is the problem of finding a shortest non-zero vector in L. Of course this problem depends on our choice of norm in $\mathbb{R}^n$. It is known that for the L_∞-norm (the max-norm) the shortest vector problem is **NP**-hard (see for instance [14]), which makes it quite unlikely that there is an efficient algorithm to find a shortest vector with respect to that norm. In Section 2 we are interested in the L_2-norm (the ordinary Euclidean norm). For the L_2-norm the shortest vector problem is still open, i.e., it is unknown whether the problem is **NP**-hard or allows a polynomial-time solution (see [3] for an algorithm that runs in polynomial time if the dimension of the lattice is fixed).

In Section 2 we have a weaker version of the shortest vector problem: it suffices to find a non-zero vector that is longer than a shortest vector by at most a factor C, for some $C > 1$. This problem can be solved as follows. Let $L = L(b_1, b_2, \ldots, b_k)$ be as above a lattice of dimension k in $\mathbb{R}^n$. In 1981 L. LovÁsz invented an algorithm, the basis reduction algorithm (see [7, Section 1]), that transforms the basis $b_1, b_2, \ldots, b_k$ for L into a *reduced* basis $\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_k$ for L. Roughly speaking, a reduced basis is a basis that is *nearly orthogonal* ; for a precise definition of this concept, and for a description of the basis reduction algorithm, we refer to [7, Section 1].

It is intuitively clear that a basis that is nearly orthogonal contains a vector that is not much longer than a shortest vector in the lattice. For a reduced basis $\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_k$ for L the following can be proved:

$$|\tilde{b}_1|^2 \leqslant 2^{k-1} \cdot |x|^2$$

for every non-zero x in L. This implies that the first vector $\tilde{b}_1$ in the reduced basis is longer than a shortest non-zero vector in L by at most a factor $2^{(k-1)/2}$. In Section 2 it is therefore sufficient to take $C = 2^{m/2}$.

In [7] it is shown that the running time of the basis reduction algorithm, when applied to a basis $b_1, b_2, \ldots, b_k$ in $\mathbb{Z}^n$, is bounded by a polynomial function of k, n, and $\max_i(\log|b_i|)$. Combined with a precise analysis of the results from Section 2 it follows that a primitive polynomial f in $\mathbb{Z}[X]$ of degree n can be factored in time polynomial in n and $\log|f|$.

Except for a polynomial-time algorithm for factoring polynomials, there exist many more applications of L. Lovász' basis reduction algorithm. To mention a few: simultaneous diophantine approximation [7], breaking knapsack based cryptosystems [1, 8], and the disproof of the Mertens conjecture [10].

REFERENCES

1. E. BRICKELL. Breaking iterated knapsacks. *Proceedings Crypto 84.*
2. G.E. COLLINS. Factoring univariate polynomials in polynomial average time. *Proceedings Eurosam 79,* 317-329.
3. R. KANNAN (1983). Improved algorithms for integer programming and

related problems. *Proceedings 15th STOC.*

4. R. KANNAN, A.K. LENSTRA, L. LOVÁSZ (1984). Polynomial factorization and nonrandomness of bits of algebraic and some transcendental numbers. *Proceedings 16th STOC.*
5. E. KALTOFEN, D.R. MUSSER, B.D. SAUNDERS (1981). A generalized class of polynomials that are hard to factor. *Proceedings ACM Symposium on Symbolic and Algebraic Computation,* 188-194.
6. D.E. KNUTH (1981). *The Art of Computer Programming, Vol. 2, Second Edition, Seminumerical Algorithms,* Reading, Addison-Wesley.
7. A.K. LENSTRA, H.W. LENSTRA, Jr., L. LOVÁSZ (1982). Factoring polynomials with rational coefficients. *Math. Ann. 261,* 515-534.
8. J.C. LAGARIAS, A.M. ODLYZKO (1983). Solving low-density subset sum problems. *Proceedings 24th FOCS.*
9. M. MIGNOTTE (1974). An inequality about factors of polynomials. *Math. Comp. 28,* 1153-1157.
10. A.M. ODLYZKO, H. TE RIELE (1985). Disproof of the Mertens conjecture. *J. reine und angew. Math. 357.*
11. A. SCHÖNHAGE (1982). *The Fundamental Theorem of Algebra in Terms of Computational Complexity,* Preliminary Report, Math. Inst. Univ. Tübingen.
12. A. SCHÖNHAGE (1984). Factorization of univariate integer polynomials by diophantine approximation and an improved basis reduction algorithm. *Proceedings 11th ICALP, LNCS* 172, 436-447.
13. B.L. VAN DER WAERDEN (1931). *Moderne Algebra,* Springer, Berlin.
14. P. VAN EMDE BOAS (1981). *Another* **NP**-*complete Partition Problem and the Complexity of Computing Short Vectors in a Lattice,* Report, Univ. Amsterdam.
15. H. ZASSENHAUS (1969). On Hensel factorization, I. *J. of Number Theory 1,* 291-311.