WHAT YOUR EIGHTH GRADE TEACHER NEVER TOLD YOU ABOUT FACTORING

Edmund A. Lamagna
Department of Computer Science
University of Rhode Island
Kingston, RI 02881 USA

E-mail: eal@cs.uri.edu

The methods we learned in algebra to factor polynomials rely on a certain amount of pattern matching, guesswork, and just plain luck. They work well for quadratics, a few special cubics, and perhaps an occasional polynomial of higher degree where one or two roots are obvious. One of the more impressive capabilities of computer algebra systems is their ability to factor polynomials reliably and efficiently. How do these systems solve this problem?

We begin by addressing the question of what it means to factor a polynomial. Let us examine the response of a typical computer algebra system, Maple, to a few simple factorization problems.

$$\texttt{factor}(x^2-9) = (x-3)(x+3) \qquad (1) \qquad\qquad \texttt{factor}\,(x^2-7) = x^2-7 \qquad (4)$$

$$\texttt{factor}\,(4x^2+16x+16) = 4\,(x+2)^2 \qquad (2) \qquad\qquad \texttt{factor}\,(x^2+9) = x^2+9 \qquad (5)$$

$$\texttt{factor}\,(x^2-9/16) = 1/16\,(4x-3)\,(4x+3) \quad (3)$$

The first two answers are not surprising. By analogy with (1), we probably expected $(x–3/4)$ $(x+3/4)$ to be returned for (3). The polynomials in (4) and (5) are returned unfactored. Why not answer $(x-\sqrt{7})$ $(x+\sqrt{7})$ and $(x–3i)$ $(x+3i)$, respectively?

When factoring polynomials we need to do so with respect to some algebraic structure, and computer algebra systems choose to factor over the integers unless directed otherwise. Factoring polynomials with rational coefficients is equivalent to factoring over the integers. To factor

$$p(x) = a_n/b_n\, x^n + a_{n-1}/b_{n-1}\, x^{n-1} + \ldots + a_1/b_1\, x + a_0/b_0\,,$$

we can remove $1/\mathrm{lcm}(b_n,b_{n-1},\ldots,b_1,b_0)$ from $p(x)$, leaving a polynomial with integer coefficients. So factoring $x^2–9/16$ is equivalent to factoring $1/16\ (16x^2–9)$, explaining Maple's answer to (3).

1. Schubert-Kronecker Method

Finding the linear factors of a polynomial $p(x) = c_nx^n + c_{n-1}x^{n-1} +\ldots+c_1x + c_0$ is equivalent to determining its rational roots. Observe that if $x = a/b$ (reduced to lowest terms) is a root of $p(x)$ then a divides c_0 and b divides c_n.

Example 1-1. Suppose $p(x) = 6x^4–x^3+4x^2–x–2$. The divisors of $c_0 = -2$ are ±1, ±2 and those of $c_n = 6$ are ±1, ±2, ±3, ±6, and so the candidate rational roots are ±2, ±1, ±2/3, ±1/2, ±1/3, ±1/6. Substituting each into $p(x)$, we find $p(2/3) = 0$, $p(-1/2) = 0$, and $p(x) \neq 0$ for the other 10 choices. So, $(x–2/3)\ (x+1/2) = 1/6\ (3x–2)\ (2x+1)$ evenly divides $p(x)$, and $p(x) = 1/6\ (3x–2)\ (2x+1)\ (6x^2+6)$ $= (3x–2)\ (2x+1)\ (x^2+1)$. □

Isaac Newton, in his *Arithmetica Universalis* (1707), described a method for finding the linear and quadratic factors of a univariate polynomial. In 1793, the astronomer Friedrich von Schubert showed how to extend this technique to find factors of any degree. Schubert's method was rediscovered by Leopold Kronecker some 90 years later. Their procedure is fairly straightforward to understand and implement, and we now describe it.

The key ideas behind the Schubert-Kronecker method are:

1) If p can be factored, then one factor has degree $m \le \lfloor \text{degree}(p)/2 \rfloor$.
2) If $q(x)$ is a factor of $p(x)$, then $q(a)$ evenly divides $p(a)$ for any a.
3) A unique polynomial of degree d (or less) can be interpolated through d+1 points.

Here's how Schubert-Kronecker factorization works.

Step 1. Choose m+1 points: $a_0, a_1, \ldots, a_m$. (0, ±1, ±2, ... will work.)

Step 2. Evaluate $p(x)$ at each of these points, $x = a_i$.

Step 3. Find the set f_i of distinct (+ and – factors) of $p(a_i)$.

Step 4. To find the factors of $p(x)$ of degree d, choose combinations of d+1 points (a_i, b_i), where $b_i \varepsilon f_i$, and interpolate $q(x)$. If $q(x)$ divides $p(x)$, then $q(x)$ is a factor of $p(x)$.

Example 1-2. Suppose we want to factor $p(x) = x^4+4$. $m = \lfloor \text{degree}(p)/2 \rfloor = 2$, so let's choose the m+1 points: $a_0 = 0$, $a_1 = 1$, $a_2 = -1$. Evaluating p at these points and factoring the values, we have

$$\begin{array}{ll} p(a_0) = p(0) = 4 & f_0 = \{\pm1, \pm2, \pm4\} \\ p(a_1) = p(1) = 5 & f_1 = \{\pm1, \pm5\} \\ p(a_2) = p(-1) = 5 & f_2 = \{\pm1, \pm5\} \end{array}$$

p has no linear factors. To find its quadratic factors, there are 6·4·4 = 96 triples to test. One triple that doesn't succeed is $b_0 = 1$, $b_1 = -1$, $b_2 = 5$. Interpolating $q(x) = ax^2+bx+c$, we obtain

$$\begin{array}{lrcl} (0,1): & 1 & = & a\cdot0^2 + b\cdot0 + c \\ (1,-1): & -1 & = & a\cdot1^2 + b\cdot1 + c \\ (-1,5): & 5 & = & a\cdot(-1)^2 + b\cdot(-1) + c \end{array}$$

Solving, we find that $a = 1$, $b = -3$, $c = 1$, and $q(x) = x^2-3x+1$ does not divide $p(x)$. A triple that works is $b_0 = 2$, $b_1 = 5$, $b_2 = 1$. Here we have

$$\begin{array}{lrcl} (0,2): & 2 & = & a\cdot0^2 + b\cdot0 + c \\ (1,5): & 5 & = & a\cdot1^2 + b\cdot1 + c \\ (-1,1): & 1 & = & a\cdot(-1)^2 + b\cdot(-1) + c \end{array}$$

and $a = 1$, $b = 2$, $c = 2$. This time $q(x) = x^2+2x+2$ evenly divides $p(x)$ and the quotient is x^2-2x+2. Therefore, $p(x) = (x^2+2x+2)\,(x^2-2x+2)$. □

Unfortunately, the running time of the Schubert-Kronecker algorithm is exponential in the degree of p, and so it is practical only for polynomials of very low degree.

2. Simplifying the Problem

Our problem can be simplified greatly by restricting the types of polynomials considered. We first show how the factorization of a polynomial with an arbitrary leading coefficient can be determined from that of an "equivalent" monic polynomial.

Suppose we want to factor a polynomial $p(x)$ whose leading coefficient $c_n \neq 1$. We perform the following steps:

Step 1. Form $q(x)$ by substituting x/c_n for x in $p(x)$ and multiplying by c_n^{n-1}.
Step 2. Factor $q(x)$.
Step 3. Reverse the substitution: $p(x) = 1/c_n^{n-1}\, q(c_n x)$.

Example 2-1. Here's a simple example, $p(x) = 2x^2+3x+1$. First, we form

$$q(x) \;=\; 2^1\,(2\,(x/2)^2 + 3\,(x/2) + 1) \;=\; x^2 + 3x + 2\,.$$

Next, we factor the monic polynomial: $q(x) = (x+2)\,(x+1)$. Finally, we reverse the substitution to uncover the factorization

$$p(x) \;=\; 1/2^1\,(2x + 2)\,(2x + 1) \;=\; (x + 1)\,(2x + 1)\,. \;\square$$

If we start with polynomials that are "primitive" (*i.e.*, the greatest common divisor (gcd) of the original coefficients is removed), we can remove the gcd of the coefficients from each factor produced at Step 3 to obtain the final result.

A second simplification involves the detection of repeated factors. These are found quickly by computing the *square-free decomposition.* The following theorem provides a test for repeated factors. Its proof embodies an algorithm for computing the square-free decomposition that involves taking only derivatives and polynomial greatest common divisors.

Theorem 1. p has repeated factors if and only if $\gcd(p, p') = 1$.

Example 2-2. A polynomial in square-free form is

$$p(x) \;=\; (x^2 + 1)\,(x^2 - 1)^4\,(x^3 + 3x)^5\,.$$

Once the square-free form has been found, each of its components can be factored separately to produce the complete factorization,

$$p(x) \;=\; (x^2 + 1)\,(x - 1)^4\,(x + 1)^4\,x^5\,(x^2 + 3)^5\,.$$

Note that $p'(x) = 2x\,(x^2{-}1)^4\,(x^3{+}3x)^5 + 8x\,(x^2{+}1)\,(x^2{-}1)^3\,(x^3{+}3x)^5 + 15\,(x^2{+}1)^2\,(x^2{-}1)^4\,(x^3{+}3x)^4$ and $\gcd(p, p') = (x^2 - 1)^3\,(x^3 + 3x)^4 \neq 1$. □

With the two simplifications presented in this section, our attention can be directed to the problem of factoring a monic polynomial with no repeated factors.

3. Roundabout Factorization

We noted that the Schubert-Kronecker algorithm runs very slowly when implemented on a computer. So how do today's computer algebra systems factor? They employ a "roundabout" technique first suggested by Kurt Hensel in 1908. This method, shown diagrammatically in Figure 1, is based on factoring a polynomial modulo a prime number m, and then "lifting" this result to a factorization over the integers. The key idea behind roundabout factorization is the following.

Theorem 2. If $p(x)$ is monic and $p(x) = s(x)\,t(x)$ over the integers, then $p(x) = s(x)\,t(x) \bmod m$ for every prime m.

$$\begin{tikzcd}
\text{$p(x)$ in $Z[x]$} \arrow[r, dash] \arrow[d, dash, "\text{choose prime } m"'] & \text{Factors of $p(x)$ in $Z[x]$} \\
\text{$p(x)$ mod m in $Z_m[x]$} \arrow[r, dash] & \text{Factors of $p(x)$ mod m in $Z_m[x]$} \arrow[u, dash, "\text{lift}"']
\end{tikzcd}$$

Figure 1. Roundabout Factorization

Note, importantly, the reverse of Theorem 2 is not always true! For example, x^2+1 is irreducible over the integers but factors mod 2 as $(x+1)^2$. x^4+1 is irreducible over the integers but factors modulo *every* prime. The next two examples illustrate the usefulness of the theorem.

Example 3-1. Consider $p(x) = x^6+4$ and its modulo 11 and 13 factorizations:

$$\begin{aligned} p(x) &= (x^2 + 5)\,(x^2 + 2x + 5)\,(x^2 + 9x + 5) \quad \text{mod } 11 \\ &= (x^3 + 3)\,(x^3 + 10) \quad \text{mod } 13 \end{aligned}$$

Since p has three factors of degree 2 mod 11, any factorization over the integers must involve either (1) three factors of degree 2 or (2) one factor of degree 4 and one of degree 2. Similarly, since p has two factors of degree 3 mod 13, its integer factorization must also involve two factors of degree 3. These two modular factorizations are incompatible, so p is irreducible over the integers. (This can also be inferred immediately from the fact that p is irreducible mod 7.) □

Example 3-2. Let's determine the factorization of $p(x) = x^6-1$ from its modulo 3 factorization,

$$p(x) = (x + 1)^3\,(x + 2)^3 \text{ mod } 3\,.$$

$p_1 = x+1$ mod 3 corresponds to $x+1$ over the integers, and $p_2 = x+2$ mod 3 corresponds to $x-1$. Since p_1 and p_2 evenly divide p over the integers (*i.e*, rem(p, p_1) = rem(p, p_2) = 0), they are both factors of p. After dividing p_1 and p_2 out of p, what remains is $q(x) = x^4+x^2+1$. Moreover since rem(q, p_1) = rem(q, p_2) = 3, p_1 and p_2 are not repeated factors of p over the integers. Next, we take the p_is pairs:

$$\begin{aligned} p_1 p_2 &= x^2 + 2 \text{ mod } 3 = x^2 - 1 \text{ over } Z[x] \\ p_1{}^2 &= x^2 + 2x + 1 \text{ mod } 3 = x^2 - x + 1 \text{ over } Z[x] \\ p_2{}^2 &= x^2 + x + 1 \text{ mod } 3 = x^2 + x + 1 \text{ over } Z[x] \end{aligned}$$

and see if each product divides q (and hence p) over the integers. We find that rem(q, $p_1{}^2$) = rem(q, $p_2{}^2$) = 0 but rem(q, $p1\,p2$) = 3, so $p_1{}^2$ and $p_2{}^2$ are factors of p over the integers. Therefore, the desired factorization is $p(x) = (x+1)\,(x-1)\,(x^2-x+1)\,(x^2+x+1)$. □

We had just enough "wiggle room" in Example 3-2 to deduce the factorization over the integers from the modulo 3 factorization since all the coefficients of the factored polynomial are between -1 and $+1$. We do not, in general, know the coefficients of the factored polynomial in advance. However, an inequality attributed to Maurice Mignotte (1974), building upon earlier work of Edmund G. H. Landau, enables us to get a fairly good bound on their size, thereby enabling us to choose a suitable prime modulus m.

4. Distinct Degree Factorization

At the heart of the roundabout method is the ability to factor a polynomial modulo a prime number m. Two strategies are available: one due to Elwyn R. Berlekamp (1967) and the other to David G. Cantor and Hans Zassenhaus (1979). The latter technique, called *distinct degree factorization*, is considerably simpler and we describe it here. The basis of the method is the following remarkable theorem.

Theorem 3. $x^{m^r}-x$ is the product of all irreducible polynomials over $Z_m[x]$ whose degree divides r.

Example 4-1. Let's use the theorem to find the irreducible polynomials modulo 2 of degrees 1, 2 and 3. Setting $m = 2$ and $r = 1$, we find that

$$x^2 - x \;=\; x\,(x+1) \quad \mathrm{mod}\, 2$$

is the product of all irreducible polynomials of degree 1 over $Z_2[x]$. Substituting $r = 2$,

$$x^4 - x \;=\; x\,(x+1)\,(x^2+x+1) \quad \mathrm{mod}\, 2$$

is the product of all irreducible polynomials of degrees 1 and 2. Similarly with $r = 3$,

$$x^8 - x \;=\; x\,(x+1)\,(x^3+x^2+1)\,(x^3+x+1) \quad \mathrm{mod}\, 2$$

is the product of all irreducible polynomials of degrees 1 and 3. □

Example 4-2. Now let's use Theorem 3 to factor $p(x) = x^{15}+1$ modulo 2. Letting $q_d(x) = x^{2^d}-x$, we can determine the linear factors (degree $d = 1$) by finding the greatest common divisor of p and q_1 mod 2,

$$s_1(x) \;=\; \gcd(p, q_1) \;=\; \gcd(x^{15}+1, x^2-x) \;=\; x+1 \quad \mathrm{mod}\, 2\,.$$

So p has 1 factor of degree 1, $s_1(x) = x{+}1$. Now divide s_1 out of p, then find the degree $d = 2$ factors the same way,

$$\begin{aligned} s_2(x) \;&=\; \gcd(p/s_1, q_2) \;=\; \gcd(x^{14}+x^{13}+x^{12}+x^{11}+x^{10}+x^9+x^8+x^7+x^6+x^5+x^4+x^3+x^2+x+1,\, x^4-x) \\ &=\; x^2+x+1 \quad \mathrm{mod}\, 2\,. \end{aligned}$$

There is one degree 2 factor, $s_2(x) = x^2+x+1$. Now look for any degree $d = 3$ factors,

$$s_3(x) \;=\; \gcd(p/s_1s_2, q_3) \;=\; \gcd(x^{12}+x^9+x^6+x^3+1, x^8-x) \;=\; 1 \quad \mathrm{mod}\, 2\,.$$

There are none. Next look for degree $d = 4$ factors,

$$\begin{aligned} s_4(x) \;&=\; \gcd(p/s_1s_2s_3, q_4) \;=\; \gcd(x^{12}+x^9+x^6+x^3+1, x^{16}-x) \\ &=\; x^{12}+x^9+x^6+x^3+1 \quad \mathrm{mod}\, 2\,. \end{aligned}$$

There are three degree 4 factors, and that's all that's left of p. The problem remains of splitting these three degree 4 factors,

$$s_4(x) = (x^4+x^3+x^2+x+1)\,(x^4+x^3+1)\,(x^4+x+1) \quad \mathrm{mod}\, 2\,.$$

This is done using another set of gcd computations described by Cantor and Zassenhaus. □

Bibliography

A. G. Akritas, *Elements of Computer Algebra with Applications.* Wiley: 1989.

K. O. Geddes, S. R. Czapor and G. Labahn, *Algorithms for Computer Algebra.* Kluwer: 1992.

D. E. Knuth, *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms.* Addison-Wesley: 1969, 1981, 1998.