RC 20280 (11/17/95)
Computer Science / Mathematics

IBM Research Report

Finding a Small Root of a Bivariate Integer Equation; Factoring with high bits known

Don Coppersmith

IBM Research Division
T.J. Watson Research Center
Yorktown Heights, New York

IBM Research Division
Almaden · T.J. Watson · Tokyo · Zurich

Finding a Small Root of a Bivariate Integer Equation; Factoring with high bits known

Don Coppersmith*

Abstract

We present a method to solve integer polynomial equations in two variables, provided that the solution is suitably bounded. As an application, we show how to find the factors of $N = PQ$ if we are given the high order $(1/4+\epsilon)(log N)$ bits of P. This compares with Rivest and Shamir's factor 1/3.

1 Introduction

We present a method to solve a polynomial equation $p(x, y) = 0$ over $\mathbf{Z}$, provided that the solution is suitably bounded: $|x| < X$ and $|y| < Y$, with X, Y depending on the coefficients and degree of p.

Our algorithm uses lattice basis methods [2]. It is similar in spirit to [1], which solved equations in one variable in ($\mathbf{Z}$ mod N), but the present algorithm requires a different analysis.

We require bounds X and Y on the absolute values of x and y in our solution. Suppose $p(x, y)$ has degree δ in each variable, and $p(x, y) = \sum_{ij} p_{ij} x^i y^j$. Define $D = max_{ij} |p_{ij}| X^i Y^j$ as the largest possible term in $p(x, y)$ in the region of interest. Then we will find a bounded solution (x, y) (if it exists) provided that

$$XY < D^{\{2/(3\delta)\}-\epsilon}.$$

For fixed degree δ, the algorithm runs in time polynomial in $log D$ and $1/\epsilon$.

Similar methods can be applied to the multivariate case but are not assured of success; the proof breaks down at a critical point.

Our immediate application, and the framework in which the algorithm is described, is the problem of factoring an integer when we know its high order bits. If we know $N = PQ$ and we know the high order $(1/4+\epsilon)(log N)$ bits of P, then by solving the equation $(P_0+x)(Q_0+y)-N$ over a suitable range of x and y we can find the factorization of N. By comparison, Rivest and Shamir [3] need about $(1/3)(log N)$ bits of P. This has applications to some RSA-based cryptographic schemes; see for example [4].

2 Factoring with high order bits known

We can apply the present techniques to the problem of factoring an integer when we know the high-order bits of one of the factors.

*IBM T.J. Watson Research Center, Yorktown Heights, New York, USA; copper@watson.ibm.com

Suppose we know $N = PQ$, and suppose we know the high order $(1/4 + \epsilon)(log N)$ bits of P. By division we know the high order bits of Q as well.

$$\begin{aligned}
P &= P_0 + x_0 \\
Q &= Q_0 + y_0 \\
|x_0| &< P_0/N^{1/4+\epsilon} = X \\
|y_0| &< Q_0/N^{1/4+\epsilon} = Y \\
p(x,y) &= (P_0 + x)(Q_0 + y) - N \\
&= (P_0Q_0 - N) + Q_0x + P_0y + xy \\
p(x_0, y_0) &= PQ - N = 0
\end{aligned}$$

Here P_0 and Q_0 are known, while x_0 and y_0 are unknown, and x and y are dummy variables.

We will relate the bounds X and Y to the quantity

$$D = max(|P_0Q_0 - N|, Q_0X, P_0Y, XY).$$

This is approximately the largest possible size of an individual term of $p(x, y)$ with bounded x and y. In the case of a more general polynomial $p(x, y) = \sum_{ij} p_{ij} x^i y^j$, we would define

$$D = max_{ij}(|p_{ij}|X^iY^j).$$

We are trying to find a bounded pair of integers (x, y) solving $p(x, y) = 0$. We begin by selecting an integer $k > 1/(8\epsilon)$.

For all pairs of integers (i, j) with $0 \leq i \leq k$ and $0 \leq j \leq k$, form the polynomial $q_{ij}(x, y) = x^i y^j p(x, y)$. Obviously $q_{ij}(x_0, y_0) = 0$.

Form a matrix M_1 with $(k + 2)^2$ rows, indexed by pairs (g, h) with $0 \leq g, h \leq k + 1$; and $(k+2)^2+(k+1)^2$ columns, the left-hand columns indexed by pairs (g, h) with $0 \leq g, h \leq k+1$, and the right-hand columns indexed by pairs (i, j) with $0 \leq i, j \leq k$. The left-hand block is a diagonal matrix whose $((g, h), (g, h))$ entry is given by $X^{-g}Y^{-h}$. The $((g, h), (i, j))$ entry of the right-hand block is the coefficient of $x^g y^h$ in the polynomial $q_{ij}(x, y)$.

Perform elementary row operations on M_1 to produce a matrix M_2 whose right-hand block has the $(k + 1)^2$ identity matrix on the bottom and the $(2k + 3)$ by $(k + 1)^2$ zero matrix on the top. We can do this because the coefficient of xy in p is 1, so that the right-hand block of M_1 contains an upper triangular matrix with 1 on the diagonal.

The lattice formed by these top $(2k + 3)$ rows of M_2 is the sublattice of the original lattice gotten by setting to 0 all the right-hand columns.

Now do lattice basis reduction on the top $2k + 3$ rows of M_2; let the resulting $2k + 3$ rows form a new matrix M_3.

Consider the $(k + 2)^2$-long row vector $\mathbf{r}$ whose (g, h) entry is $x_0^g y_0^h$. The row vector $\mathbf{s}$ of length $(k + 2)^2 + (k + 1)^2$ given by $\mathbf{s} = \mathbf{r}M_1$ satisfies

$$\begin{aligned}
\mathbf{s}_{(g,h)} &= (x_0/X)^g (y_0/Y)^h \\
|\mathbf{s}_{(g,h)}| &\leq 1 \\
\mathbf{s}_{(i,j)} &= q_{ij}(x_0, y_0) = 0 \\
|\mathbf{s}| &< k + 2
\end{aligned}$$

Because its right-hand side is 0, $\mathbf{s}$ is one of the vectors in the row lattice spanned by M_3. We will show that it is a "relatively short" vector in the lattice, which will enable us to confine it to a hyperplane, thus producing a linear equation relating its coefficients. This will translate directly to a polynomial equation on x and y which is not a multiple of $p(x, y)$. We can then take the resultant to find a single polynomial equation satisfied by x_0, and solve this equation over **Z**.

We proceed to estimate the sizes of the vectors in M_3. Define the diagonal matrix W of dimension $(k+2)^2 \times (k+2)^2$, with $((g,h),(g,h))$ entry given by $X^g Y^h$. In the matrix WM_1, the left-hand block is the identity. In the (i,j) column of the right-hand side of WM_1, the largest element has absolute value $X^i Y^j D$. Further, the right-hand columns are "nearly orthogonal", because they are part of a Toeplitz array.

So there is a choice of $(k+1)^2$ columns in the left-hand block whose deletion (transforming a rectangular matrix M to a square matrix $\hat{M}$) leaves a matrix $W\hat{M}_1$ with determinant

$$\begin{aligned} det(W\hat{M}_1) &= \Omega(\textstyle\prod_{ij}(X^i Y^j D)) \\ &= \Omega((XY)^{(k+1)^2 k/2} D^{(k+1)^2}), \end{aligned}$$

the constant implicit in Ω depending on k and the pattern of nonzero coefficients in the polynomial $p(x, y)$. (For this particular polynomial the constant is 1.) (The proof is probably standard, and not particularly enlightening or relevant. I will sketch it in the appendix.)

For this particular polynomial the constant is 1 (see appendix) so we can drop the "Ω".

Since

$$det(W) = \prod_{gh}(X^g Y^h) = (XY)^{(k+2)^2(k+1)/2},$$

we can calculate

$$det(\hat{M}_1) \geq D^{(k+1)^2}(XY)^{-(k+1)(3k+4)/2} = (D^{k+1}(XY)^{-(3k+4)/2})^{k+1}.$$

From

$$D \geq |Q_0 X| = Q_0 P_0 / N^{1/4+\epsilon}$$

$$XY = (P_0/N^{1/4+\epsilon})(Q_0/N^{1/4+\epsilon}) = P_0 Q_0 / N^{1/2+2\epsilon}$$

we see

$$D^{k+1}(XY)^{-(3k+4)/2} \geq (P_0 Q_0)^{-(k/2)-1} N^{(k/2)+(3/4)+(2k+3)\epsilon}.$$

Because $P_0 Q_0 = N(1 + o(N^{-1/4}))$ we can replace $P_0 Q_0$ by N and incur negligible error:

$$D^{k+1}(XY)^{-(3k+4)/2} \geq N^{(-1/4)+(2k+3)\epsilon}(1 + o(kN^{-1/4})).$$

Recall $k > 1/(8\epsilon)$. Then

$$D^{k+1}(XY)^{-(3k+4)/2} > N^{3\epsilon}(1 + o(k^2 N^{-1/4})),$$

$$det(\hat{M}_1) \geq N^{(3\epsilon)(k+1)}(1 + o(k^2 N^{-1/4})) \geq N^{3/8} >> 1.$$

The same estimate applies to $det(\hat{M}_2)$ and to the determinant of the left-hand $2k+3$ square submatrix of $\hat{M}_3$. For the following discussion, we call that square submatrix L and its dimension $n = 2k+3$. From the discussion in [2], the reduced basis $b_1, ..., b_n$ of L satisfies

$$|b_n^*| \geq det(L)^{1/n}/2^{(n-1)/4},$$

where b_n^* is the component of b_n orthogonal to the subspace spanned by $b_1, ..., b_{n-1}$. If we have chosen k sufficiently large, this will be larger than $|\mathbf{s}|$:

$$|b_n^*| = \Omega(N^{(3/8)(1/(2k+3))}/2^{(2k+2)/4}) > k + 2.$$

Suppose this holds. Then for any row vector $\mathbf{t}$ in the lattice spanned by L, if $\mathbf{t}$ is not in the lattice spanned by $b_1, ..., b_{n-1}$, then its expression as an integer combination of the b_i involves b_n nontrivially. Thus we have $|\mathbf{t}| \geq |b_n^*| > k + 2 > |\mathbf{s}|$. But for $\mathbf{s}$ itself, the $n = 2k + 3$ entries corresponding to those columns in the left-hand side that remain when we transform M to $\hat{M}$ form a row vector $\hat{\mathbf{s}}$ in the lattice spanned by L (since the right-hand elements of $\mathbf{s}$ are 0). Also, $|\hat{\mathbf{s}}| \leq |\mathbf{s}|$. Thus $\hat{\mathbf{s}}$ is in the lattice spanned by $b_1, ..., b_{n-1}$.

Membership in this subspace gives us a linear relation on the coefficients $x_0^g y_0^h$ expressing $\mathbf{s}$ as a linear combination of the rows of M_1. This relation is linearly independent of the $(k+1)^2$ relations given by the polynomials $q_{ij}(x, y)$ which determined that $\mathbf{s}$ had right-hand side 0 and thus was in the lattice of M_3 to start with. So this relation translates to a polynomial relation $q(x_0, y_0) = 0$ where $q(x, y)$ is not a polynomial multiple of $p(x, y)$.

Take the resultant of $p(x, y)$ and $q(x, y)$ with respect to y to discover a nontrivial integer polynomial $r(x)$ in one variable x satisfied by x_0: $r(x_0) = 0$. Since q has degree at most $k + 1$ in each variable and p has degree 1 in each variable, r has degree at most $2k + 2$. Solve $r(x_0) = 0$ in the reals to find a small number of candidates for x_0, among which the correct value can be gotten by inspection.

3 Complexity

We had to invoke a lattice basis reduction on a matrix of size $2k+3 \approx 1/(4\epsilon)$, whose elements were integers of bit-length $O(logN)$.

4 Discussion on lattice methods

The lattice basis reduction method is inherently linear. If we want to relate several unknowns by a polynomial equation $p(x, y) = \sum_{gh} p_{gh} x^g y^h = 0$, one natural approach is to replace each monomial $x^g y^h$ by a new independent variable z_{gh}, and let p become a linear relation among several independent bounded variables: $\sum_{gh} p_{gh} z_{gh} = 0$, $|z_{gh}| \leq X^g Y^h = Z_{gh}$. In order to get results, we would need something like $\prod Z_{gh} \leq D$, and this would imply severe restrictions on X and Y.

In the present paper we have extended this approach by using several polynomials q_{ij} but reusing the same independent bounded variables z_{ij}. We are able to amortize the cost of the several variables over several equations. This accounts for the success of the present method in increasing the feasible sizes of X and Y.

5 Extension to other polynomials; more variables

Similar techniques could be applied to other polynomial equations besides the given $p(x_0, y_0) = 0$. They are less effective for polynomials of more than two variables; see below.

Even in the case of two variables, the present technique is sensitive to the form of the polynomial $p(x,y)$. For an arbitrary quadratic polynomial we could not tolerate ranges X, Y for the unknowns x, y as large as we did here. This is because although $p(x,y)$ has degree two, it does not have terms involving x^2 or y^2, only xy.

When we estimated $det(\hat{M}_1)$ by dividing the estimate of $det(W\hat{M}_1)$ by $det(W)$, there was considerable cancellation in the powers X and Y. The term $det(W\hat{M}_1)$ had powers X^iY^j for each pair (i,j) with $0 \le i, j \le k$; these represented the monomials x^iy^j by which $p(x,y)$ was multiplied. The term $det(W)$ had powers X^gY^h for each pair (g,h) with $0 \le g, h \le k+1$; these represented the monomials x^gy^h appearing in these products $x^iy^jp(x,y)$. The range on g exceeds the range on i by 1 because p has degree 1 in x. If p had an x^2 term, we would have needed to enlarge the range of g. The powers of X and Y appearing in the ratio $det(W\hat{M}_1)/det(W)$ arise from the pairs (g,h) outside the range of (i,j), namely

$$\{(g,h) | g = k+1, 0 \le h \le k\} \cup \{(g,h) | h = k+1, 0 \le g \le k+1\}.$$

The inclusion of an x^2 term would have enlarged that region by another layer. This would have led directly to a stricter requirement on the sizes of X and Y.

Preliminary estimates are that if $p(x,y)$ has degree G in x and H in y, then we can tolerate ranges X and Y satisfying

$$D >> X^{G+(H/2)}Y^{H+(G/2)}$$

by allowing $0 \le i, j \le k$, or more generally about

$$D >> X^{G+(\alpha H/2)}Y^{H+(G/(2\alpha))}$$

by allowing $0 \le i \le k\alpha$ and $0 \le j \le k$ for any positive α.

If $p(x,y)$ has total degree G then we can tolerate about

$$D >> (XY)^G$$

by allowing pairs (i,j) with $i \ge 0$, $j \ge 0$, and $i + j \le k$. This is better than the previous approach if p is a general bivariate polynomial of total degree G, but worse if (like the current example) it is really of degree $G/2$ separately in each variable.

If we have a polynomial $p(x,y,z)$ in three variables, we can mimic the present approach. If the ranges X, Y, Z are small enough, we will end up with a polynomial relation $q(x,y,z)$, not a multiple of p, satisfied by (x_0, y_0, z_0). Then the resultant of p and q with respect to z will give a polynomial $r(x,y)$ in two variables. We can then solve r by the current methods. But the degree of r will be quite high, so that the ranges X and Y which can be tolerated will be quite small.

We still have a heuristic procedure which *might* work for a multivariate polynomial. We are guaranteed to find a space of codimension 1 (a hyperplane) containing all the short vectors of the lattice M_3. But we might easily find a space of larger codimension. (There is a good possibility that many basis vectors $|b_i^*|$ exceed our known upper bound on $|\mathbf{s}|$, and each one increases the codimension of the space which contains all the short vectors.) We develop several polynomial equations $q_i(x,y,z)$ satisfying $q_i(x_0,y_0,z_0) = 0$; the number of such equations is equal to the codimension of this space. We can then take resultants and gcds of the various q_i and p and hope to produce a single polynomial equation in a single variable $r(x_0) = 0$, which we solve over the reals. This is only a heuristic approach, which might or might not work for a given polynomial p.

References

[1] Don Coppersmith, "Finding a Small Root of a Univariate Modular Equation," IBM RC 20223, 10/11/95, revised 11/08/95.

[2] A. K. Lenstra, H. W. Lenstra and L. Lovasz, "Factoring Polynomials with Integer Coefficients," *Matematische Annalen* vol. 261 (1982), 513-534.

[3] R. L. Rivest and A. Shamir, "Efficient factoring based on partial information," *Advances in Cryptology - EUROCRYPT '85,* Lecture Notes in Computer Science, Vol. 219, Berlin: Springer-Verlag, 1986, pages 31-34.

[4] S. A. Vanstone and R. J. Zuccherato, "Short RSA Keys and Their Generation," *Journal of Cryptology* volume 8 number 2 (Spring 1995), 101-114.

6 Appendix

We need to show that several columns of a Toeplitz matrix are "nearly orthogonal". That is, given a fixed column vector V with L entries, the largest of which (in absolute value) is D, and given a matrix M whose K columns are copies of V shifted by different offsets and padded with 0, we can find a $K \times K$ submatrix of M with determinant at least $D^K 4^{-KL^2}$ in absolute value.

Let $V(i), i = 1, 2, ..., L$ be complex numbers, and

$$D = max_i |V(i)| = |V(I)|.$$

Define $V(i) = 0$ if $i < 1$ or $i > L$. Let $c_1, ..., c_K$ be integer offsets, $0 \le c_1 < c_2 < ... < c_K$. Define a matrix M of size $(L + c_K) \times K$, with

$$M_{ij} = V(i - c_j).$$

We show how to find a $K \times K$ submatrix of M with a relatively large determinant.

Find an index H such that

$$|V(H) 4^{(H-I)^2}| = max_j |V(j) 4^{(j-I)^2}| \ge D.$$

Select the rows $H + c_i$ of M to create the desired $K \times K$ submatrix $\hat{M}$. Then $\hat{M}_{ij} = V(H + c_i - c_j)$.

Multiply the ith row of $\hat{M}$ by $4^{2(H-I)c_i}$ and the jth column by $4^{-2(H-I)c_j}$. This does not change the determinant. The ij entry is now

$$V(H + c_i - c_j) 4^{2(H-I)(c_i - c_j)}.$$

By our choice of H we have

$$|V(H + c_i - c_j)| 4^{(H + c_i - c_j - I)^2} \le |V(H)| 4^{(H-I)^2}.$$

So the magnitude of the ij entry is bounded by

$$|V(H)| 4^{(H-I)^2 - (H - I + c_i - c_j)^2 + 2(H-I)(c_i - c_j)} = |V(H)| 4^{-(c_i - c_j)^2} \le |V(H)| 4^{-(i-j)^2},$$

while the diagonal entries are all equal to $V(H)$. Our matrix is diagonally dominant – the magnitudes of the off-diagonal entries of each row sum to less than $(2/3)|V(H)|$ – so each of its eigenvalues exceeds $(1/3)|V(H)|$ in absolute value, and the absolute value of its determinant exceeds

$$\left(\frac{1}{3}|V(H)|\right)^K \geq \left(\frac{1}{3}D4^{-(L-1)^2}\right)^K \geq D^K 4^{-KL^2}$$

This is sufficient for our purposes, since we can arrange our Toeplitz matrix such that K and L are polynomially related to k and the degree of p, and are independent of the coefficients of p.

In the particular case $p(x,y) = A + Bx + Cy + xy$, we can actually find a $K \times K$ submatrix with determinant D^K, no matter which of the four terms A, BX, CY, XY is the largest. For example, suppose BX is the largest of the four terms. Assign the index (g,h) to the row numbered $g(k+2) - h + (k+1)$. Then $BX = D$ is the last of the nonzero entries in the vector V under this numbering, and the matrix $\hat{M}$ is upper triangular with D on the diagonal. In general such a lower triangular construction will be possible whenever the maximal term $D = |p_{ij}|X^iY^j$ corresponds to a point (i,j) in the plane which is a vertex of the convex hull of all the points $(i\prime, j\prime)$ for which $p_{i\prime j\prime} \neq 0$. This happens to be true of all four terms in our particular polynomial p because each of (0,0), (0,1), (1,0), (1,1) is a vertex of that convex hull.