

A MONTE CARLO METHOD FOR FACTORIZATION

J. M. POLLARD

Abstract.

We describe briefly a novel factorization method involving probabilistic ideas.

1. Introduction.

We point out a simple method by which, apparently, a prime factor p of a number can usually be found in $O(p^{\frac{1}{2}})$ arithmetical operations, as opposed to the $O(p)$ operations required by “trial division” (e.g. [1]). The theoretical possibility of doing this was shown previously [2] in a much more complicated manner (which, however, enabled us to reach a certain precise conclusion). Our method was suggested by the ideas of [3], pp. 7–8, 25, but also has connections with [2].

Consider a sequence such as $x_0 = 2$,

$$(1) \quad x_{i+1} \equiv x_i^2 - 1 \pmod{n},$$

where n is the number we are attempting to factorize. Other polynomials of degree ≥ 2 and other starting values can be used. We generate in turn the triples

$$(2) \quad (x_i, x_{2i}, Q_i), \quad i = 1, 2, \dots,$$

where

$$(3) \quad Q_i \equiv \prod_{j=1}^i (x_{2j} - x_j) \pmod{n}.$$

Each triple is obtained from its predecessor by three applications of (1) and one multiplication in (3); thus the work involved is substantially that for four multiplications \pmod{n} . We use only four multi-length variables, those for x_i , x_{2i} , Q_i and n . Whenever i is a multiple of some number m (say, $m = 100$), we compute the greatest common divisor

$$(4) \quad d_i = \gcd(Q_i, n),$$

by one of the well-known methods [3].

If $1 < d_i < n$ then we have obtained a partial factorization of n as

$n = d_i \times (n/d_i)$. Here d_i may be composite and, if so, must be factorized by some other means (in particular, the product of the smallest factors of n will be found at the first calculation of (4), if they have not already been removed). Then we continue with modulus $n' = n/d_i$ if this number is composite and not divisible by the prime factors already found. We stop on reaching some preset maximum number of steps S , a multiple of m (say, $S = 10^4$); if the final pair (x_i, x_{2i}) is saved, we have the possibility of continuing the computation at a later date.

2. Theory.

To obtain the "theory" of this method, consider (1) with n replaced by p , a prime. The sequence is ultimately periodic, that is, there are integers $c \geq 1$ and $t \geq 0$ such that $x_0, x_1, \dots, x_{c+t-1}$ are all distinct (mod p), but that $x_{c+i} \equiv x_i \pmod{p}$ for $i \geq t$ (my previous name for the method was the " ϱ -method"; here the ϱ must be drawn starting at the bottom). Define also r as the least positive integer with $x_r \equiv x_{2r} \pmod{p}$, that is,

$$\begin{aligned} t \leq r < t+c, \quad r &\equiv 0 \pmod{c}, & \text{if } t > 0, \\ r = c, & & \text{if } t = 0. \end{aligned}$$

The function $r = r(p)$ determines how soon our algorithm will find the prime factor p of n : for after $r(p)$ steps we shall have $Q_i \equiv 0 \pmod{p}$, and the factor p will then be found at the next calculation of (4) (perhaps the product of several prime factors, with nearly the same values of $r(p)$, will be found instead).

Knowing of no other way to proceed, I make the assumption that (1) (to modulus p) constitutes a "random mapping" of the residues (mod p) in the sense of [3], p. 8. Then $c(p)$ and $t(p)$ are random variables with expectations close to

$$(5) \quad \sqrt{(\pi p/8)} = 0.6267\sqrt{p},$$

and the expectation of $c(p) + t(p)$ is close to twice this value. The expectation of $r(p)$ (not in [3]) can be shown to be close to

$$(6) \quad \pi^{5/2}\sqrt{p}/(12\sqrt{2}) = 1.0308\sqrt{p},$$

(the error terms in (5) and (6) are $O(1)$ as $p \rightarrow \infty$).

Thus we expect the mean values of $c(p)/\sqrt{p}$, $t(p)/\sqrt{p}$ and $r(p)/\sqrt{p}$ to be close to the constants in (5), (5) and (6) respectively. For the 100 largest primes below 10^6 these values were found to be 0.6127, 0.6821 and 1.0780.

We are interested also in the distribution of the values of $r(p)/\sqrt{p}$; in this direction I estimate that $r(p) < \frac{1}{2}\sqrt{p}$ with probability 0.183, and