PREFACE

- I plan to tell you about how large numbers are factored on computers. I would like you to be able to read current research papers with considerable understanding. To do research in this area, you need either to be very lucky or to understand a fair lot of number theory. (Though no more than could fit in a one-term course.)

PREREQUISITES

- Please understand some simple facts about divisibility, including greatest common divisors.
- Please understand the definition and a few simple properties of abstract groups. In particular, the order of (the number of elements in) a subgroup is a divisor of the order of the group.

- Please understand modular arithmetic and a few simple facts about it. The non-zero integers modulo a prime form a group under multiplication which is in fact a cyclic group. The integers modulo a composite number n which are relatively prime to n also form a group, which is generally not cyclic.
- Please understand Fermat's (little) theorem, which says in effect that when p is prime, then a^(p-1) is congruent to 1 modulo p. I suppose that a is not divisible by p.

HOW DOES ONE FACTOR A NUMBER

- My first comment is that the difficulty of factoring a number n does not depend primarily on the size of the number n, but rather on the size of the factor (usually prime factor) that you are looking for.
- A good general rule is that the difficulty of factoring a number n depends on the size of the second largest prime factor.

- Some of the more advanced techniques work poorly if the number n has small factors and may indeed fail if the number being factored is either an even number or is a perfect power of an integer.
- More than that, factorization methods may grind away for a long time trying to factor a number which is in fact prime, so it is well to settle that question early in the process.

TRIAL DIVISION

- By far the fastest method of finding small factors is trial division. We could always find a factorization of n by dividing n by all integers between 2 and n-1. This requires no theorems, but goes directly to the definition. The time required would be proportional to n (nearly).
- In fact, you only need to try division up to the square root of n, and now the time required is more nearly proportional to sqrt(n). More than that, you only need to try division by primes up to sqrt(n), though what is gained is offset by the time to recognize a potential divisor as prime or by the storage required to store a table of primes.

- It is a matter of judgement how far you go in fact with trial division. I have found it best to cull out all the prime factors up to about six digits.
- It is very easy to omit the numbers divisible by 2 and by 3 by only looking at potential divisors that are congruent to +1 or -1 modulo 6.
- Determining that n is not a perfect power can be done in a variety of ways, all fast. Logarithms work, as do a variety of iteration methods. They are all fast.

POLLARD

- One of the most innovative (and surprising) methods to appear in this century was developed by J. M. Pollard (with substantial contributions by Richard P. Brent).
- This description of the method is accurate, but not the usual description. Choose a number k at random between 1 and n (the number to be factored). Find the greatest common divisor of k and n. If this greatest common divisor is 1 (n and k are relatively prime), then choose another random k. If the greatest common divisor is not 1, then you have found a factor.

- Is is possible that such a method could be fast enough to be useful? Not only is it possible, it is in fact considerably faster than trial division. The number of steps required is (asymptotically) a constant multiple of sqrt(p) to find the prime factor p.
- The method does not find factors in strictly increasing order and the resulting factors may not be prime. Things work better if all of the small factors are gone.

- I generally use Pollard's p method to find prime factors up to about ten digits.
- Pollard also devised some other methods, called the p-1 and the p+1 methods, which are of considerable practical and theoretical interest. They have given rise to one of the fastest modern methods, called the number field sieve.

THE RECOGNITION OF PRIMES

- It would be nice if the converse of Fermat's theorem were true so that for a number relatively prime to n, a^(n-1) were congruent to 1 modulo n if and only if n were prime. Unfortunately , this converse is not true.
- When n is not a prime (When n is composite), then a^(n-1) may or not be congruent to 1 modulo n, but it usually is not. The exceptions are rare and are called composite pseudo primes to the base a. This exponentiation is very rapid.

- The smallest composite pseudoprimes to the base 2 are 341 and 561. There are no others less than 1000.
- After running this test for some value of a, the result is either that n is (certainly) composite or that it is probably prime.
- Of course, you can run the test for different values of a, getting more and more confidence that n is prime, but never certainty. There are a few numbers called Carmichael numbers, that are pseudoprimes to every base.

- There is also a stronger test for pseudo-primality, which involves finding successive square roots of a^(p-1) which is easy to do, and checking that no numbers (modulo n) appear other than 1 and -1. If any other number appears, then n is definitely composite, since in a cyclic group, the number 1 can have no more than two square roots. Otherwise, the number is called a strong pseudo prime.

- It is strange that a number can be proven composite with no notion of what the factors are.
- The problem has by now been solved in the sense that an intricate but fasts method can settle the question of whether n is composite or prime.

GAUSS'S ATTEMPT

- In his book, Disquisitiones Arithmeticae, Gauss tried very hard to come up with practical methods of factorization. While his attempt can scarcely be termed a failure, it was certainly not a success. It has given rise to some modern published research, but never in the forefront of methods.
- To understand this approach, you have to understand a bit more number theory than I am assuming. In particular, you need to know something about class field theory, about quadratic number fields, and about binary quadratic forms. (These are really one subject matter, not three).

FERMAT/LEGENDRE

- It was recognized early that if a number n can be represented in the form n=a^2-b^2 where a and b are integers, then n is composite and has the factors a+b and a-b.
- This method is cold comfort, since the individual operations are not speedy and the chances of hitting a perfect square by this method are poor.

- Here is an improved approach. Start with a number x near to the square root of n and form x^2-n. Attempt to factor x^2-n. If it is a perfect square, then you are done. If not, then look at the factorization. As an example, I will choose 1001. What I get (what I in fact got) were the following factorizations.

  - 65^2-n = 3224 = 2*2*2*13*31
  - 63^2-n = 2968 = 2*2*2*7*53
  - 61^2-n = 2720 = 2*2*2*2*2*5*17
  - 59^2-n = 2480 = 2*2*2*2*5*31
  - 49^2-n = 1400 = 2*2*2*5*5*7
  - 28^2-n = -617 = (-1)*7*31
  - 27^2-n = -272 = (-1)*2*2*2*2*17
  - 24^2-n = -425 = (-1)*5*5*17

Bibliography

- Hans Riesel, "Prime Numbers and Computer Methods for Factorization", Second Edition, Birkhaeuser, 1994. ISBN 0-8176-3743-5 / 3-7643-3743-5
- C. F. Gauss, "Disquisitiones Arithmeticae", Yale University Press, 1966