

A Search for Large Twin Prime Pairs

By R. E. Crandall and M. A. Penk

Abstract. Two methods are discussed for finding large integers m such that $m - 1$ and $m + 1$ are both primes. Eight such numbers m of magnitudes 22, 22, 32, 64, 136, 154, 203, and 303 digits are listed; together with primitive roots (for $m + 1$) and Lucas-Lehmer parameters (for $m - 1$). The Hardy-Littlewood twin prime conjecture is supported by a statistical test involving the generation of 249 twin prime pairs in the 50-to-54 digit region.

1. Asymptotic Estimates. We define a twin prime mean (TPM) to be any positive integer m such that $(m - 1, m + 1)$ is a pair of primes. It is unknown whether $M(x)$, the number of TPMs not exceeding x , is bounded as x increases. However, in 1919 V. Brun [1], by showing in effect that for some positive constant C ,

$$M(x) < \frac{Cx}{\log^2 x} (\log \log x)^2$$

established that the sum of the reciprocals of all TPMs must converge. Twin prime pairs $(m - 1, m + 1)$ are, therefore, significantly sparser than the primes themselves, so the task of discovering large TPMs is both interesting and challenging.

For the computer search reported here, we have made practical use of certain asymptotic estimates concerning the distribution of TPMs. Evidently, if m is a TPM, then for any prime $p < m - 1$, we have $m^2 \not\equiv 1 \pmod{p}$. In particular, if $m > 4$, then $m \equiv 0 \pmod{6}$. We define a sieving set with sieve limit x to be the set of all primes p with $5 \leq p \leq x$; this set is denoted as P_x . Then a function $Q(x)$ is defined

$$Q(x) = \prod_{p \in P_x} \left(1 - \frac{2}{p}\right) \sim \frac{K}{\log^2 x},$$

where K is a positive constant. For an integer $n \equiv 0 \pmod{6}$, much larger than x but otherwise "random", we loosely interpret $Q(x)$ as the "probability" that $n^2 - 1$ is prime to all primes in P_x . In 1923 Hardy and Littlewood used functions similar to Q to derive an heuristic estimate [2]

$$M(z) \sim \frac{2C_2 z}{\log^2 z},$$

where C_2 is a constant $0.66016181\dots$ related to K and Euler's constant γ by $C_2 = Ke^{2\gamma}/4$. The Hardy-Littlewood estimate is apparently in good agreement with actual counts of TPMs [3], [4], [5].

Received March 31, 1978; revised June 15, 1978.

AMS (MOS) subject classifications (1970). Primary 10A25, 10A40.

Key words and phrases. Prime, twin primes, Hardy-Littlewood conjecture.

© 1979 American Mathematical Society
0025-5718/79/0000-0028/\$02.50

To test whether some $m \equiv 0 \pmod{6}$ is a TPM it is natural to first choose a sieve limit x and test both $m + 1$ and $m - 1$ for divisors in P_x . In the absence of any such divisors we then apply Fermat tests, that is to determine whether $2^m \equiv 1 \pmod{m + 1}$ and $2^{m-2} \equiv 1 \pmod{m - 1}$ as must both be true if $m \pm 1$ are both primes. In practice, these tests are virtually conclusive; and we assume in computer time calculations that almost all pairs $m \pm 1$ which pass the Fermat tests are twin prime pairs. From these considerations the mean time to locate a TPM can be estimated from $Q(x)$, which is used to derive sieving times and the frequency of required Fermat tests; and from the Hardy-Littlewood estimate, which estimates the probability of a successful pair of Fermat tests. The problem of determining the most efficient sieve limit x is, thus, reduced to a straightforward minimization problem involving the relevant asymptotic formulas.

2. Primality Tests. Once a possible TPM m is located using the sieve and Fermat tests we can test $m + 1$ for true primality as follows. If b is an integer such that $b^m \equiv 1 \pmod{m + 1}$ but for every prime factor q of m we have $b^{m/q} \not\equiv 1 \pmod{m + 1}$, then b is a primitive root of the prime $m + 1$ [5].

The number $m - 1$ can be tested using a Lucas-Lehmer test [7]. Let Q be chosen so that $D = 1 - 4Q$ satisfies $D^{(m-2)/2} \equiv -1 \pmod{m - 1}$ and $\text{g.c.d.}(m - 1, QD) = 1$. Define $U_0 = 0$, $U_1 = 1$ and for $n \geq 2$, $U_n = U_{n-1} - QU_{n-2}$. If $U_m \equiv 0 \pmod{m - 1}$ but for every prime factor q of m we have $U_{m/q} \not\equiv 0 \pmod{m - 1}$, then $m - 1$ is prime. We call Q a Lucas-Lehmer parameter for the prime $m - 1$.

For these primality tests to be directly applicable it is necessary to know the prime factors of m . In the present search a useful factorization algorithm has been Pollard's rho method [8], [9]. The method was used, for example, to find the two ten-digit factors of the second TPM listed in Table 1.

The battery of computer algorithms used in this search is comprised of:

- (1) sieve algorithm, sieve limit x ;
- (2) power algorithm $b^y \pmod{z}$;
- (3) Lucas-Lehmer test $U_k \pmod{z}$, parameter Q ;
- (4) Pollard's rho algorithm.

Fermat testing is of, course, a special case of (2), and is assumed to be a process of order $O(\log^3 z)$.

3. Method of Interval-Sieving. This search method finds in principle all TPMs in an interval, though at the expense of necessitating factorization of any suspected TPMs. Choose some large integer $n \equiv 0 \pmod{6}$ as the first member of an arithmetic progression $\{n, n + 6, \dots, n + 6(u - 1)\}$. We establish a binary array of bits $b(1), \dots, b(u)$ all set initially to zero and choose a sieve limit x . For each prime $p \in P_x$ we compute $r^\pm(n, p)$ as the least positive values of $\pm 6^{-1} - 1 + (n/6) \pmod{p}$. Then, we set bits $b(r^\pm(n, p) + Lp) = 1$ for all nonnegative L such that the array argument does not exceed u . The result of this sieve operation is that the only possible TPMs in the progression are of the form $m = n + 6(j - 1)$ where $b(j) = 0$. Approximately

$u \cdot Q(x)$ numbers m will so survive the sieve, and these m are to be Fermat-tested.

With a sieve limit $x = 2000$ the two 22-digit TPMs in Table 1 were found. There are no TPMs lying strictly between these two.

With sieve limit $x = 10000$ the 32-digit TPM of the table was found. For numbers of this size and smaller, a small sieve and Pollard rho method in conjunction are generally sufficient for factorization of m . On the idea that perhaps 60-digit factorizations are beyond the reach of present techniques [8], it is clear that an alternative search method is required for TPMs having, say, hundreds of digits.

4. A Monte Carlo Method. One of us (M.P.) has refined a search method which circumvents factorization altogether. The idea is to establish an a priori form

$$m = 2 \cdot 3 \cdot s_1 \cdots s_j \cdot q_1 \cdots q_k,$$

where the s_i are (possibly) manually chosen small primes, and the primes q_i are generated randomly subject to the constraints that $q_i < 2^W$, where W is a machine arithmetic word length; and m is in some predetermined region.

The procedure is to choose a sieve limit x and possible factors s_i ; then to perform these steps:

- (1) generate the q_i and form m ;
- (2) sieve $m \pm 1$ with all primes in P_x ;
- (3) Fermat-test $m + 1$;
- (4) Fermat-test $m - 1$.

On failure at any step, one reverts back to step (1). When a trial m survives all four steps it is a likely TPM, and true primality tests are to follow.

The procedure was programmed on a PDP 11/70 computer with the outer loops written in the language *C* and the inner loops in assembly language. The word length of the machine is $W = 15$; and the basic arithmetic process, a W -by- W bit multiply and two adds, consumed 12 microseconds.

A formula was derived for the expected time to find an m which survives all four steps in terms of the rough magnitude of m and the sieve limit x . In the case that no s_i are chosen, the formula arises as follows. Enumerate the consecutive primes in P_x as p_1, \dots, p_n and denote for convenience $Q(p_0) = 1$. The probability that p_k is the smallest prime dividing $m^2 - 1$ is taken to be $(2/p_k)Q(p_{k-1})$. An estimate for the expectation of the number of sieve divisions at step (2), per choice of m in step (1), is

$$\sum_{k=1}^n \frac{2k}{p_k} Q(p_{k-1}) \sim \frac{Ax}{\log^4 x},$$

where A is a constant. The actual time spent for a sieve division of $m \pm 1$ is assumed to be proportional to $\log m$ but essentially independent of k , so that the expected time spent at step (2), per choice of m in step (1), is estimated as

$$\frac{Bx \log m}{\log^4 x},$$

for a machine-dependent constant B . Since a fraction $Q(x)$ of the m chosen in step (1) will pass step (2), the expected time spent at step (3), per choice of m , is estimated as

$$CQ(x)\log^3 m,$$

where C is another machine constant. The time spent at step (4) is assumed to be negligible relative to that spent at step (3). From the Hardy-Littlewood formula we expect $O(\log^2 m)$ choices of m in step (1) before we find a number which passes all four steps. Such a number will be found, therefore, in an expected time

$$T = \frac{Hx \log^3 m}{\log^4 x} + \frac{L \log^5 m}{\log^2 x},$$

where H, L are machine constants.

A good choice for x which essentially minimizes T is that for which

$$\pi(x) = (2L/H)\log^2 m.$$

This implies that the number of primes in the most efficient sieving set P_x is proportional to the square of the number of digits in m .

With the computer actually used for the search, the ratio L/H was empirically determined to be 0.02, implying that searching in, say, the 100-digit region should involve an x with $\pi(x)$ about equal to 2100. Thus, the choice $x = 2^{15}$, $\pi(x) = 3512$, an extremely convenient one for the programmer, turned out to be an efficient one as well. The 64, 136, 154, 203, and 303 digit TPMs in Table 1 were found with this method. Though $x = 2^{15}$ is highly efficient only for 100 to 150 digits the same value for x was maintained in the searches for the largest TPMs with only a moderate degradation of efficiency.

5. A Statistical Test in the 50-54 Digit Region. A test of the Hardy-Littlewood formula was carried out using 132947 trial values for m in the method of the last section. These values were constrained by $s_1 = 5, s_2 = 7, s_3 = 11, s_4 = 13, q_i < 2^{15}$, and $10^{49} < m < 10^{54}$.

The Hardy-Littlewood formula predicts that such a computer run should produce about

$$\left(\frac{2 \cdot 3}{Q(13)}\right) \cdot 132947 \cdot \frac{2C_2}{\log^2 z}$$

TPMs, where z is some real number between 10^{49} and 10^{54} . The correction factor $(2 \cdot 3)/Q(13)$ arises from the special form for the trial m values. The correction factor for the less specialized q_i constraint is assumed to be negligible. Assuming Poisson statistics for the experimental occurrence of TPMs, this prediction amounts to 245 ± 25 TPMs, with the error corresponding to the ambiguity in z combined with one standard deviation $(245)^{1/2}$.

The results of the actual computer run are as follows. Exactly 249 numbers m passed all four steps of the Monte Carlo procedure. Each one of these m was shown to be a TPM in subsequent primality tests. The Monte Carlo procedure consumed a total of five hours C.P.U. time, and the required 498 primality tests took another five hours.

It is of interest that this statistical test gives an experimental value for C_2 which is within 2% of the Hardy-Littlewood prediction.

	digits	b	Q
$m = 2 \cdot 2 \cdot 2 \cdot 3 \cdot 3 \cdot 3 \cdot 19 \cdot 23 \cdot 487 \cdot 117998390827927$	22	7	-4
$m = 2 \cdot 3 \cdot 5 \cdot 17 \cdot 1630555079 \cdot 6522807839$	22	3	-13
$m = 2 \cdot 3 \cdot 5 \cdot 31 \cdot 61 \cdot 197947 \cdot 1077403318801219090697$	32	3	-10
$m = 2 \cdot 2 \cdot 3 \cdot 3 \cdot 7 \cdot 1453 \cdot 2969 \cdot 5779 \cdot 7589 \cdot 7603 \cdot 10039 \cdot 10937 \cdot 11887 \cdot 12239 \cdot 20297 \cdot 24103 \cdot 25583 \cdot 26237 \cdot 26423 \cdot 31513$	64	2	-3
$m = 2 \cdot 3 \cdot 3 \cdot 11 \cdot 179 \cdot 953 \cdot 1223 \cdot 1481 \cdot 2791 \cdot 4153 \cdot 6197 \cdot 6217 \cdot 8101 \cdot 8219 \cdot 10627 \cdot 11161 \cdot 12841 \cdot 13099 \cdot 16493 \cdot 16993 \cdot 17027 \cdot 17489 \cdot 17863 \cdot 19433 \cdot 19463 \cdot 20681 \cdot 21031 \cdot 21859 \cdot 22051 \cdot 23549 \cdot 27277 \cdot 27961 \cdot 28549 \cdot 28631 \cdot 28909 \cdot 29077 \cdot 32561$	136	2	-28
$m = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 29 \cdot 101 \cdot 1543 \cdot 3917 \cdot 4889 \cdot 5011 \cdot 5261 \cdot 5503 \cdot 6521 \cdot 9833 \cdot 11657 \cdot 12241 \cdot 13109 \cdot 13291 \cdot 13441 \cdot 13523 \cdot 14891 \cdot 16183 \cdot 17599 \cdot 17659 \cdot 17789 \cdot 19421 \cdot 19699 \cdot 19709 \cdot 20063 \cdot 21157 \cdot 21503 \cdot 21523 \cdot 23993 \cdot 24043 \cdot 25469 \cdot 27631 \cdot 28181 \cdot 28909 \cdot 29633 \cdot 30223 \cdot 32479$	154	2	-17
$m = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 631 \cdot 691 \cdot 1193 \cdot 1367 \cdot 1451 \cdot 1931 \cdot 1999 \cdot 3257 \cdot 3539 \cdot 4703 \cdot 4967 \cdot 4987 \cdot 5113 \cdot 5639 \cdot 6067 \cdot 6949 \cdot 7451 \cdot 8363 \cdot 8689 \cdot 9161 \cdot 9887 \cdot 10181 \cdot 10501 \cdot 11171 \cdot 11471 \cdot 11701 \cdot 11821 \cdot 12541 \cdot 13421 \cdot 13709 \cdot 14341 \cdot 16007 \cdot 17317 \cdot 17863 \cdot 18587 \cdot 19031 \cdot 19379 \cdot 20773 \cdot 21283 \cdot 25243 \cdot 25819 \cdot 26669 \cdot 27737 \cdot 27751 \cdot 28513 \cdot 29209 \cdot 29327 \cdot 29611 \cdot 30803 \cdot 30983$	203	6	-32
$m = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 241 \cdot 433 \cdot 773 \cdot 811 \cdot 1301 \cdot 1301 \cdot 1493 \cdot 2383 \cdot 4007 \cdot 4327 \cdot 4409 \cdot 4597 \cdot 4783 \cdot 5431 \cdot 5519 \cdot 5581 \cdot 5851 \cdot 5987 \cdot 6679 \cdot 7297 \cdot 7481 \cdot 7589 \cdot 7817 \cdot 8737 \cdot 8999 \cdot 10061 \cdot 10433 \cdot 10663 \cdot 11117 \cdot 11243 \cdot 11677 \cdot 11807 \cdot 13121 \cdot 13553 \cdot 14479 \cdot 14771 \cdot 15031 \cdot 15101 \cdot 15149 \cdot 15199 \cdot 15607 \cdot 16231 \cdot 16607 \cdot 16979 \cdot 17053 \cdot 17417 \cdot 17881 \cdot 18523 \cdot 19013 \cdot 19087 \cdot 19531 \cdot 21713 \cdot 21997 \cdot 22153 \cdot 23071 \cdot 23143 \cdot 24659 \cdot 24733 \cdot 24821 \cdot 25367 \cdot 26267 \cdot 26573 \cdot 26759 \cdot 27803 \cdot 27893 \cdot 28001 \cdot 28433 \cdot 29411 \cdot 29537 \cdot 29753 \cdot 29837 \cdot 30649 \cdot 31849 \cdot 31981$	303	3	-14

TABLE 1.

Twin primes $m \pm 1$ with primitive roots b for $m + 1$; and Lucas-Lehmer parameters Q for $m - 1$. All factors listed are primes. The first three m values were found using interval-sieving, so these factorizations required implementation of factoring algorithms. For the remaining five m values, the factors were chosen a priori as explained in the text.

Acknowledgements. The authors wish to thank the administration of Reed College, in whose Computer Center much of this work was performed.

2234 S. E. 20th Avenue
Portland, Oregon 97214

2133 S. E. Clinton Street
Portland, Oregon 97202

2. G. H. HARDY & J. E. LITTLEWOOD, "Some problems of Partitio Numerorum," *Acta Math.*, v. 44, 1923, pp. 1–70.
3. D. H. LEHMER, "Tables concerning the distribution of primes up to 37 millions," *MTAC*, v. 13, 1959, pp. 56–57, UMT 3.
4. G. H. HARDY & E. M. WRIGHT, *An Introduction to the Theory of Numbers*, 4th ed., Clarendon Press, Oxford, 1965.
5. R. P. BRENT, "Irregularities in the distribution of primes and twin primes," *Math. Comp.*, v. 29, 1975, pp. 43–56.
6. D. H. LEHMER, "Computer technology applied to the theory of numbers," *Studies in Number Theory*, vol. 6, W. J. LeVeque (Ed.), Math. Assoc. Amer.; distributed by Prentice-Hall, Englewood Cliffs, N. J., 1969, pp. 117–151.
7. D. E. KNUTH, *The Art of Computer Programming*, vol. 2, Addison-Wesley, Reading, Mass., 1969; Section 4.3.1, 4.5.4.
8. R. GUY, "How to factor a number," *Proc. 5th Manitoba Conference on Numerical Mathematics*, Oct. 1–4, 1975, pp. 49–89.
9. J. M. POLLARD, "A Monte Carlo method for factorization," *BIT*, v. 15, 1975, pp. 331–334.