Annals of Mathematics **00** (XXXX), 1–6
https://doi.org/10.4007/annals.XXXX.00.0.0000

A naive approach to prime factorization

By Shashankaditya Upadhyay

Abstract

A method to find prime factors of a given arbitrary number n is proposed. When n is semiprime, the method is shown to run in $O(n^2 log(log(n)))$ time. Eigenvalue distribution of matrices utilized in the proposed method is conjectured.

Contents

1. Introduction

Prime factorization has been a topic of academic interest for a long time, particularly because of its applications to cryptosystems [ROW94]. It has been shown that prime factorization for a given number is achievable in polynomial time [AKS04] and there is an ample body of work utilizing ideas originating from the fields of algebraic number theory, elliptic curves and more recently quantum computations to achieve prime factorization of a given number [Lan94, Sho99, Cha12].

It is commonly understood that finding prime factors of a given number is a difficult problem, especially for semiprimes i.e. a number which is product

Keywords: Prime factorization, Prime indicator matrices, Semiprimes
AMS Classification: Primary: 11A51 (matsc2020), 11C20 (matsc2020); Secondary: 68GW (matsc2020).
The class was commissioned by Annals of Mathematics.

of exactly two prime numbers of similar sizes and this fact forms the basis of RSA cryptosystems [RSA78, IAAA21]. However, it is possible to enlist all the prime numbers less than a given number in polynomial time using methods such as sieve of Eratosthenes [Hor72] while certain derivatives of this method are known to be computationally more efficient [Pri83, Hel20], achieving time complexity of $O(n)$, for a given number n.

In this article we make use of the fact that all the prime numbers below a certain number can be known in polynomial time to device a method that can compute prime factors of a given number, which is particularly more efficient in case of semiprimes. Moreover, for semiprimes, the proposed method is shown to run in polynomial time. The method is termed naive because it is fairly simple to understand and easy to execute.

2. Review of some known results

The prime number theorem for arithmetic progression was conjectured by Dirichlet [LD38] and was later proved by Poussin [DLVP96]. It can be stated as follows:

THEOREM 2.1. *Let a and d be positive coprime integers i.e., the greatest common divisor of a and d is 1. Let $\pi_{d,a}(x)$ denote the number of primes less than x in the arithmatic progression $a, a+d, a+2d, \ldots$, then $\pi_{d,a}(x)$ is estimated as*

$$\pi_{d,a}(x) \approx \frac{Li(x)}{\varphi(d)},$$

where $Li(x)$ is the Eulerian logarithmic integral function defined as $Li(x) = \int_2^x \frac{dt}{ln(t)}$ and φ represents the Euler's totient function which for a given positive integer d counts the number of positive integers that are coprime to d.

It is further known that the Eulerian integral function $Li(x)$ can be estimated as $Li(x) \approx \pi(x)$, where $\pi(x)$ is the prime counting function giving the number of primes less than x. This result was published posthumously by Gauss [Gau74] and later Skewes showed that $Li(x)$ and $\pi(x)$ change sign an infinite number of times [Ske33]. The prime counting function can be further approximated as $\pi(x) \approx \frac{x}{ln(x)}$, as may be familiar to most readers [DLVP96].

In a later section, we make use of the prime number theorem and approximations for the functions $Li(x)$ and $\pi(x)$ to prove the main result of this work. However, for the time being, we would like to introduce the notion of prime indicator matrices.

3. Prime indicator matrices of order n

Let n be a given prime or composite number. Consider an $n \times n$ square matrix $A_n = (a_{ij})$ such that for $1 \leq i, j \leq n$, $a_{ij} = (j-1) * n + i$. Using A_n,

we can define a matrix $P_n = (b_{ij})$, such that $b_{ij} = 1$, if a_{ij} is a prime number and $b_{ij} = 0$ otherwise. For the purpose of this work we can call such a matrix P_n as prime indicator matrix of order n. In particular, the author is unaware whether such matrices exist in the literature and have been studied in the past.

For the purpose of computation, a prime indicator matrix of order n can be computed in $O(n^2)$ time [Pri83] (in $O(n^2 log(log(n)))$ time using sieve of Eratosthenes [Hor72]), without specifically finding the matrix A_n.

For example, the prime indicator matrix of order 6 can be seen as

$$P_6 = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

4. Prime factorization using prime indicator matrices

Let $n \geq 9$ be a given number and let the prime decomposition of n be equal to ${p_1}^{m_1}{p_2}^{m_2} \dots {p_k}^{m_k}$, where $k \geq 1$ and $m_1, m_2 \dots m_k$ are the multiplicities of the prime factors $p_1, p_2 \dots p_n$ respectively. If n is a prime number, then $k = 1$ and $m_k = m_1 = 1$. If n is semiprime then $k = 2$ and $m_1 = m_2 = 1$.

The prime decomposition of a given number n using the prime indicator matrix of order n is derived from the following theorem.

THEOREM 4.1. *Let $n \geq 9$ be a given number such that the prime decomposition of n is equal to ${p_1}^{m_1}{p_2}^{m_2} \dots {p_k}^{m_k}$ and let P_n be the prime indicator matrix of order n. The row sum of the p^{th} row of the matrix P_n, where p is a prime, is equal to 1 if and only if $p|n$.*

Proof. Starting with the converse part, we can argue that when $p|n$, such that p is prime, then $p \in \{p_1, p_2 \dots p_k\}$ and the entry in the first column of the p^{th} row of matrix P_n is 1 while the rest of the entries in the p^{th} row are 0 since the corresponding entries in matrix A_n are of the form $r * n + p_i$ where $p_i = p \in \{p_1, p_2 \dots p_k\}$ and $1 \leq r \leq n - 1$. This is so because $r * n + p_i$ is the same as $r * ({p_1}^{m_1}{p_2}^{m_2} \dots {p_k}^{m_k}) + p_i$, which is equal to $p_i * [r * ({p_1}^{m_1}{p_2}^{m_2} \dots {p_{i-1}}^{m_{i-1}}{p_i}^{m_i - 1}{p_{i+1}}^{m_{i+1}} \dots {p_k}^{m_k}) + 1]$ and thus each such entry is a composite number.

For the if part, i.e., when the row sum of the p^{th} row (p is prime) of the matrix P_n is exactly equal to 1, we can have two cases: either $p \in \{p_1, p_2 \dots p_k\}$ or $p \notin \{p_1, p_2 \dots p_k\}$, i.e., p is coprime to n. When $p \in \{p_1, p_2 \dots p_k\}$, there is not much to prove as in this case again the entry in the first column of the p^{th} row of matrix P_n is 1 while the rest of the entries in the p^{th} row have to be 0

since the corresponding entries in matrix A_n are of the form $r * n + p$ where $p = p_i \in \{p_1, p_2 \dots p_k\}$ and $1 \leq r \leq n-1$.

For the case where $p \notin \{p_1, p_2 \dots p_k\}$, that is, when p is coprime to n, one can notice that the entries in the p^{th} row of the corresponding matrix A_n are $p, p+n, p+2n \dots p+(n-1)n$ and thus form an arithmetic progression with p and n as coprime. Thus in this case, the prime number theorem for arithmetic progression can be applied to find the number of primes less than or equal to $n(n-1)+p$. Thus the number of primes appearing in the arithmetic progression can be estimated as:

$$\pi_{n,p}(n(n-1)+p) \approx \frac{Li((n-1)n+p)}{\varphi(n)} \geq \frac{Li((n-1)n)}{\varphi(n)} \approx \frac{\pi((n-1)n)}{\varphi(n)} \approx \frac{(n-1)n}{ln((n-1)n)\varphi(n)}.$$

Since the largest value $\varphi(x)$ can take is $x-1$, when x is a prime, we have

$$\frac{(n-1)n}{ln((n-1)n)\varphi(n)} \geq \frac{(n-1)n}{ln((n-1)n)*(n-1)} = \frac{n}{ln((n-1)n)}.$$

It can be easily verified that $\frac{n}{ln((n-1)n)} > 2$ for all $n \geq 9$. Thus the prime number theorem for arithmetic progression guarantees that there are at least two prime numbers in p^{th} row of matrix A_n for $n \geq 9$. As a result, when $p \notin \{p_1, p_2 \dots p_k\}$ the row sum of the p^{th} row of the prime indicator matrix P_n is at least 2. Thus for $n \geq 9$, if row sum of the p^{th} is 1 then p should belong to the set $\{p_1, p_2 \dots p_k\}$. Thus $p|n$.

□

We are now in a position to derive methods to find prime decomposition of a given number n by employing the theorem proved above. We can safely assume that prime indicator matrix P_n for the given integer n can be computed in $O(n^2 log(log(n)))$ time. When n is an arbitrary composite number of the form $n = {p_1}^{m_1} {p_2}^{m_2} \dots {p_k}^{m_k}$ then all the factors of n can be found by traversing down the first column of the corresponding prime indicator matrix P_n and taking a row sum each time 1 is encountered. Since there are n entries to be checked in the first column and row sum is taken whenever a prime is encountered, the procedure is expected to run in not more than $O(n^2)$ time. Once all the prime factors are found, we can find the multiplicities of the primes by finding the inverse of the prime factors and multiplying with the number, both of which are at most $O(n)$ operations, where inverse is calculated exactly once and multiplication is carried out m_i times for each prime factor p_i, $1 \leq i \leq k$.

Likewise, when n is a semiprime then both the prime factors can be found in $O(n^2)$ time by traversing down the first column and comparing the row sum of each row with 1 if 1 is encountered as a value in the first column. Thus forming the prime indicator matrix acts as a bottleneck for the current problem and thus overall time complexity of finding factors of a semiprime number n turns out to be at most $O(n^2 log(log(n)))$.

Moreover, if n is a prime number then only the last row of the corresponding prime indicator matrix will have a row sum of 1 and thus it can be established that n is prime in $O(n^2)$ time (given that P_n is known). For the purpose of computations, a faster strategy of course in each of the cases above would be to terminate the row sum calculation as soon as the sum becomes greater than 1. Thus for purpose of computations one need to calculate only the first k, where k is a small number compared to n columns of the prime indicator matrices so that the overall complexity of computation of prime factors becomes n.

5. **Eigenvalues of a prime indicator matrix of order** n

In this final section, we conjecture without proof that for $n \geq 4$, the eigenvalues of a prime indicator matrix of order n are distributed as:

$$-(\tfrac{n^2}{6} - \tfrac{2n}{36} - \tfrac{1}{3}), 0 \ldots 0, [\tfrac{n(n+1)}{2} + (\tfrac{n^2}{6} - \tfrac{2n}{36} - \tfrac{1}{3})].$$

References

[AKS04] M. Agrawal, N. Kayal, and N. Saxena, Primes is in p, *Annals of mathematics* (2004), 781–793.

[Cha12] K. Chandrasekharan, *Elliptic functions*, **281**, Springer Science & Business Media, 2012.

[DLVP96] C. De La Vallee-Poussin, Recherches analytiques sur la théorie des nombres premiers, *Ann. Soc. Sc. Bruxelles* (1896).

[Gau74] C. F. Gauss, *Carl Friedrich Gauss Werke*: 6, Koniglichen Gesellschaft der wissenschaften, 1874.

[Hel20] H. Helfgott, An improved sieve of eratosthenes, *Mathematics of Computation* **89** no. 321 (2020), 333–350.

[Hor72] S. Horsley, Ko kinon epato eno . or, the sieve of eratosthenes. being an account of his method of finding all the prime numbers, by the rev. samuel horsley, f. r. s., *Philosophical Transactions* (1683-1775) **62** (1772), 327–347. Available at http://www.jstor.org/stable/106053.

[IAAA21] R. Imam, Q. M. Areeb, A. Alturki, and F. Anwer, Systematic and critical review of rsa based public key cryptographic schemes: Past and present status, *IEEE access* **9** (2021), 155949–155976.

[Lan94] S. Lang, *Algebraic number theory*, **110**, Springer Science & Business Media, 1994.

[LD38] G. Lejeune Dirichlet, Sur l'usage des séries infinies dans la théorie des nombres., *Journal für die reine und angewandte Mathematik* (*Crelles Journal*) **1838** no. 18 (1838), 259–274.

[Pri83] P. Pritchard, Fast compact prime number sieves (among others), *Journal of algorithms* **4** no. 4 (1983), 332–344.

[ROW94] H. Riesel, J. Oesterlé, and A. Weinstein, *Prime numbers and computer methods for factorization*, **2**, Springer, 1994.

[RSA78] R. L. Rivest, A. Shamir, and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, *Communications of the ACM* **21** no. 2 (1978), 120–126.

[Sho99] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM review* **41** no. 2 (1999), 303–332.

[Ske33] S. Skewes, On the difference π (x)—li (x)(i), *Journal of the London Mathematical Society* **1** no. 4 (1933), 277–283.

Shrimadh Ganagativardhan Dynamics and Logistics Research Labs, 419, C Block, Sector XU-II, Geeater Noida, Gautam Buddha Nagar, U.P. 201310, India

E-mail: su348@snu.edu.in

ORCID: 0009-0009-0120-6573